

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2003, 07:34:56 ; Search time 42 Seconds
(without alignments)
452.323 Million cell updates/sec

Title: US-09-941-831A-20
Perfect score: 2343
Sequence: 1 MRESGKLDIPISDFGRMGI.....QAINQLDIPPLMISYLKQL 449

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/prodata/2/1aa/5A-COMB.pep: *
2: /cgn2_6/prodata/2/1aa/5B-COMB.pep: *
3: /cgn2_6/prodata/2/1aa/6A-COMB.pep: *
4: /cgn2_6/prodata/2/1aa/6B-COMB.pep: *
5: /cgn2_6/prodata/2/1aa/6CTUS-COMB.pep: *
6: /cgn2_6/prodata/2/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	480	20.5	350	4	US-09-302-769-25
2	293	12.5	458	4	US-09-811-469-10
3	293	12.5	458	4	US-09-811-469-11
4	264	11.3	621	3	US-09-121-964-8
5	264	11.3	621	4	US-09-155-078-2
6	219	9.3	1166	4	US-09-350-982C-5
7	209	8.9	949	3	US-09-196-387-10
8	209	8.9	949	4	US-09-841-835-10
9	208.5	8.9	1327	3	US-09-196-387-2
10	208.5	8.9	1327	4	US-09-841-835-2
11	202.5	8.6	1839	2	US-09-172-977-4
12	202	8.6	1088	3	US-09-082-059-2
13	201.5	8.6	303	2	US-09-031-485-23
14	201.5	8.6	303	2	US-08-847-429A-23
15	201.5	8.6	303	3	US-09-065-474-23
16	201.5	8.6	303	4	US-09-557-034-23
17	201.5	8.6	352	3	US-09-065-474-139
18	201.5	8.6	352	4	US-09-557-034-139
19	201.5	8.6	1745	2	US-09-031-485-33
20	201.5	8.6	1745	2	US-08-847-429A-33
21	201.5	8.6	1745	3	US-09-065-474-33
22	201.5	8.6	1745	3	US-09-557-034-33
23	198	8.5	843	2	US-09-172-977-3
24	195	8.3	1619	4	US-09-392-812A-4
25	195	8.3	1704	4	US-09-392-812A-2
26	194.5	8.3	302	2	US-09-031-485-38
27	194.5	8.3	302	2	US-08-847-429A-38

28	194.5	8.3	302	3	US-09-065-474-38	Sequence 38, Appl
29	194.5	8.3	302	4	US-09-557-034-38	Sequence 38, Appl
30	191.5	8.2	673	3	US-09-196-387-8	Sequence 8, Appl
31	191.5	8.2	673	4	US-09-841-835-8	Sequence 8, Appl
32	176.5	7.5	1709	4	US-09-392-812A-6	Sequence 6, Appl
33	168.5	7.2	679	2	US-08-462-481-4	Sequence 4, Appl
34	168.5	7.2	679	2	US-08-436-771-6	Sequence 6, Appl
35	168.5	7.2	679	2	US-08-434-998-6	Sequence 6, Appl
36	168.5	7.2	679	2	US-08-487-797-6	Sequence 6, Appl
37	168.5	7.2	679	2	US-08-701-005A-4	Sequence 4, Appl
38	168.5	7.2	679	2	US-08-473-895-4	Sequence 4, Appl
39	168.5	7.2	679	3	US-08-943-956A-4	Sequence 4, Appl
40	168.5	7.2	679	5	PCT-US95-02058-6	Sequence 6, Appl
41	164	7.0	191	2	US-09-031-485-20	Sequence 20, Appl
42	164	7.0	191	2	US-08-847-429A-20	Sequence 20, Appl
43	164	7.0	191	3	US-09-065-474-20	Sequence 20, Appl
44	164	7.0	191	4	US-09-557-034-20	Sequence 20, Appl
45	161.5	6.9	348	2	US-09-031-485-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-09-302-769-25
; Sequence 25, Application US/09302769
; Patent No. 6323317
; GENERAL INFORMATION:
; APPLICANT: HILTON, Douglas J
; APPLICANT: ALEXANDER, Warren S
; APPLICANT: VINEY, Elizabeth M
; APPLICANT: WILLSON, Tracey A
; APPLICANT: RICHARDSON, Rachael T
; APPLICANT: STARR, Robyn
; APPLICANT: NICHOLSON, Sandra E
; APPLICANT: METCALF, Donald
; APPLICANT: NICOLA, Nicos A
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
; FILE REFERENCE: 10976Z
; CURRENT APPLICATION NUMBER: US/09/302,769
; CURRENT FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 08/962,560
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Mouse
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (167)
; OTHER INFORMATION: Xaa is unsure
US-09-302-769-25

Query Match	20.5%	Score 480;	DB 4;	Length 350;
Best Local Similarity	38.3%	Pred. No. 8.4e-45;		
Matches 114;	Conservative 58;	Mismatches 104;	Indels 22;	Gaps 8;
QY	170	LNQAVYDNDSTLDQLLRQERYKRFINGRSW---	GVPGTPLRLAASYGHLSCLOVLLAH	226
Db	56	LHDAAYVGLQTLRNLLOEESYRSINEKSVCCGWLCTPLRIATAAGHNCVDFLIRK	115	
QY	227	GADVDSLVKQATPLFTAVSHGHLDVCVRLLEAGASPGSIYNNCSPLVTAARDGAVAIL	286	
Db	116	GAEVDLVGVKQTALYVAVVNGHLESTELLFAGADPNGRHRRSTPPVYHAXRVGRDDIL	175	
QY	287	QELLDHGAENVKAKL-----PVWASNTAS---	CSGPLYLAAYGHLDCFRLLHLLHGDGP	338
Db	176	KALIRYGADVDDVNNHNSDTRPPFSRLTSLVVC--	PLYISAYHNLCQFRLLQAGAMP	233
QY	339	DYNCCTDQGLLARVPR--PRTLLLEICLNHCEPEYIQLIDFGANIYLV---	PSLSLDLTQSQ	393

```
Db 234 DFNCGPNTQEFYRGSPGWCMDAFLRHGCEAAAFVSLVVEFGANLNLVKWESLGPARGR 293
Qy 394 ---DDKGIALLOARATPSLLSQVLVVRALCOAQOQAINQLDIPWMLISYLKHQ 448
Db 294 RKMDPALQVFKEARIPRTLLSLCRVARRALGKY-RLHLVPSLPDPPIKKFLLEYE 350

RESULT 2
US-09-811-469-10
; Sequence 10, Application US/09811469
; Patent No. 6551809
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; FILE REFERENCE: CL001171
; CURRENT FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Human
US-09-811-469-10

Query Match 12.5%; Score 293; DB 4; Length 458;
Best Local Similarity 27.7%; Pred. No. 1.1e-23;
Matches 86; Conservative 49; Mismatches 91; Indels 84; Gaps 13;

Qy 3 ESKWKLIPIIDFGRMGIPNRNWTITDANRNYEICSTYPPPEIVVPKSVTLGTGVGSSKER 62
Db 129 ENGKLYDPLLEYYRQGGIPNESWRITKINERYELCDTPALLVVPANIPDEELKRVASPR 188
Qy 63 SKERPVLISYKKNNAACRCQSPLSGFY-TRCVDELLELLEAISQTNPGSQFMVYVDTR 121
Db 189 SRGRIPVLSWIHPESQATITRCSPQVMVGSGKSKDEKYLQAIMDSNAQSHKIFIDAR 248
Qy 122 PKI-----WHFLVIMRIVLQAKNLM--DITKIFSLLOPKDEE--- 159
Db 249 PSNNAVANKAGGYESDAYQNAELVF---LDIHNHVMRESLRKLEIVYVNIETHW 305
Qy 160 ----EDT-----DTEKQALNQAVYDNDY-----TLDQLLRQERYK 192
Db 306 LSNLESTHWLEHIKILAGALRIADKVESGKTSVVVHCSGDGWDRTAQLTSLAMLMDGY 365
Qy 193 RFINS-----RSGWGVPGTPLRLAASGYHLSCLQVLLAHG---ADVDSLKVKAOTPLFT 243
Db 366 RTIRGFEVLVEKEW-----LSFGHR--FQLRVGHGDKNHADAD-----RSPVFL 407
Qy 244 AVSHGLDCV 253
Db 408 Q----FIDCV 413

RESULT 4
US-09-121-964-8
; Sequence 8, Application US/09121964
; Patent No. 6124447
; GENERAL INFORMATION:
; APPLICANT: Ratori, Shunji
; TITLE OF INVENTION: NOVEL ENZYME CATALYZING DEPHOSPHORYLATION
; FILE REFERENCE: 32290-144753
; CURRENT APPLICATION NUMBER: US/09/121,964
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 621
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-121-964-8

Query Match 11.3%; Score 264; DB 3; Length 621;
Best Local Similarity 26.7%; Pred. No. 3.1e-20;
Matches 80; Conservative 46; Mismatches 106; Indels 69; Gaps 12;

Qy 5 GWKLIPIIDFGRMGIPNRNWTITDANRNYEICSTYPPPEIVVPKSVTLGTGVGSSKER 64
Db 181 GMTVYNPEEYRQGLPNHWRITFINKCYELCDTPALLVVPYRASDDDLRRVATFRSR 240
Qy 65 ERVPVLSYKKNNAACRCQSPLSGFY-TRCVDELLELLEAISQTNPGSQFMVYVDTRPK 123
Db 241 NRIPVLSWIHPENKTVIVRCSPQVMVGSKRNKDKDEKYLQAIMDSNAQSHKLIYDAR 300
Qy 124 I-----WHFLVIMRIVLQAKNLM--DITKIFSLLOPKDEE--- 159
Db 301 VNAVANKATGGGYESDDAYHNAELFF---LDIHNHVMRESLRKLEIVYVNIETHW 357
Qy 160 --EDTTEE--KQALNQAVYDNDY--TLDQLLRQERYK--FINSRSGVPGTFLRLA--- 211
Db 358 SLESTHWLEHIKILVUTGAIQVADK-----VSSGKSVLVHCSGDGWDRTAQLTSLAML 411
Qy 212 -----ASYGHLSCLQVLLAHGADVDSLKVKAOTPLFTAVSHGLDCV 253
Db 412 DSFYRSIEGFEILVQKEWISFGHKFASRI--GHG-DKNHTDAD--RSPIFLQ----FIDCV 463
```

```
Db 234 DFNCGPNTQEFYRGSPGWCMDAFLRHGCEAAAFVSLVVEFGANLNLVKWESLGPARGR 293
Qy 394 ---DDKGIALLOARATPSLLSQVLVVRALCOAQOQAINQLDIPWMLISYLKHQ 448
Db 294 RKMDPALQVFKEARIPRTLLSLCRVARRALGKY-RLHLVPSLPDPPIKKFLLEYE 350

RESULT 2
US-09-811-469-10
; Sequence 10, Application US/09811469
; Patent No. 6551809
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; FILE REFERENCE: CL001171
; CURRENT FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Human
US-09-811-469-10

Query Match 12.5%; Score 293; DB 4; Length 458;
Best Local Similarity 27.7%; Pred. No. 1.1e-23;
Matches 86; Conservative 49; Mismatches 91; Indels 84; Gaps 13;

Qy 3 ESKWKLIPIIDFGRMGIPNRNWTITDANRNYEICSTYPPPEIVVPKSVTLGTGVGSSKER 62
Db 129 ENGKLYDPLLEYYRQGGIPNESWRITKINERYELCDTPALLVVPANIPDEELKRVASPR 188
Qy 63 SKERPVLISYKKNNAACRCQSPLSGFY-TRCVDELLELLEAISQTNPGSQFMVYVDTR 121
Db 189 SRGRIPVLSWIHPESQATITRCSPQVMVGSGKSKDEKYLQAIMDSNAQSHKIFIDAR 248
Qy 122 PKI-----WHFLVIMRIVLQAKNLM--DITKIFSLLOPKDEE--- 159
Db 249 PSNNAVANKAGGYESDAYQNAELVF---LDIHNHVMRESLRKLEIVYVNIETHW 305
Qy 160 ----EDT-----DTEKQALNQAVYDNDY-----TLDQLLRQERYK 192
Db 306 LSNLESTHWLEHIKILAGALRIADKVESGKTSVVVHCSGDGWDRTAQLTSLAMLMDGY 365
Qy 193 RFINS-----RSGWGVPGTPLRLAASGYHLSCLQVLLAHG---ADVDSLKVKAOTPLFT 243
Db 366 RTIRGFEVLVEKEW-----LSFGHR--FQLRVGHGDKNHADAD-----RSPVFL 407
Qy 244 AVSHGLDCV 253
Db 408 Q----FIDCV 413

RESULT 3
US-09-811-469-11
; Sequence 11, Application US/09811469
; Patent No. 6551809
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; FILE REFERENCE: CL001171
; CURRENT FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 458
; TYPE: PRT
```

```

;
; NAME/KEY: misc feature
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc feature
; LOCATION: (1102)..(1102)
; OTHER INFORMATION: n is any nucleic acid
; NAME/KEY: misc feature
; LOCATION: (2650)..(2650)
; OTHER INFORMATION: n is any nucleic acid
; US-09-350-982C-5

Query Match          9.3%; Score 219; DB 4; Length 1166;
Best Local Similarity 30.3%; Pred. No. 9.7e-15;
Matches 74; Conservative 26; Mismatches 78; Indels 66; Gaps 7;

QY 189 ERYKRF-----INRSRGWGVPTPLRLAASVCHLSCLQVLLAHGADVDSLDVKAOTPLFT 243
DB 38 ERVKRLVTPKVNRSRDTAGRKSTPLHLAAGFGKRDVBYLLQNGANVOARDDGGGLIPLHN 97
QY 244 AVSHGHLCVRVLLAAGASPGGSIYNNCSPIVLTAAARDGAVAILQELLDHGAENYKAKLP 303
DB 98 ACSFGHAEVNVLLRHGADPNARDNWNVTPLHEAAIKGIDVCIVLLQHGAEPTIRNTDG 157
QY 304 VWASNIA-----SC-----SGPLYL 318
DB 158 RTALDLADPSAKAVLTGEYKDELLESARSGNEEKWMALLTPLNVNCHASDGKSTPLHL 217
QY 319 AAVYCHLDCFRLLLLHGADPDYNCCTDGLLARVPRPTLLLEICLHNC---EPEYIQLLI 375
DB 218 AAGYNRVKIVQLLQHGA--DVHAKDKGDL--VP-----LHNACSYGHYEVELLV 264
QY 376 DFGA 379
DB 265 KHGA 268

RESULT 7
US-09-196-387-10
; Sequence 10, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRP1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521

;
; NAME/KEY: misc feature
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc feature
; LOCATION: (1102)..(1102)
; OTHER INFORMATION: n is any nucleic acid
; NAME/KEY: misc feature
; LOCATION: (2650)..(2650)
; OTHER INFORMATION: n is any nucleic acid
; US-09-350-982C-5

Query Match          11.3%; Score 264; DB 4; Length 621;
Best Local Similarity 26.7%; Pred. No. 3.1e-20;
Matches 80; Conservative 46; Mismatches 106; Indels 68; Gaps 12;

QY 5 GWKLDIPISDFRGMIPNENWITDANRNYELCSYPPPIVVPKSVTLGTWVGSKFRSK 64
DB 181 GWTVPNPEEYRQGLNHNHWRITPINKCYELCDTPALLVVPYRASDDDLRRVATFRSR 240
QY 65 ERVPVLSLYKNNAAICRCSOPLSGFY--TRCVDDELLLEAISQTNPGSQFVYVVDTRPK 123
DB 241 NRIPVLSWIHPENKTVIVRCSOPLVMSGKRNDKEDYLDVIREINKQISKLTIIDARPS 300
QY 124 I-----WHFLVLMRIVLQAKNLM--DITKIFSLQPDKEB----- 159
DB 301 NNAVANKATGGYSDDAYHNAELFF---LDIHNHVMRESLKKVKDIPVNPVEEHWLS 357
QY 160 --EDTDTER--KQALNQAVYDNDSTYLDQLLRQERYKRFINSRSGWVPGTGLRLA---- 211
DB 358 SLESTHWLHKLVLGTGAIQVADK-----VSGSKSVLVHCSGWDGMDRTAQLTSLAML 411
QY 212 -----ASYGHLSCLQVLLAHGADVDSLDVKAOTPLFTAVSHGHLDVCV 253
DB 412 DSFYRSIEGFELLVQKEWISFGHKFASRI--GHG-DKNHTDAD-RSPIFLQ----FIDCV 463

RESULT 6
US-09-350-982C-5
; Sequence 5, Application US/09350982C
; Patent No. 6455290
; GENERAL INFORMATION:
; APPLICANT: Berthelsen, Jens
; APPLICANT: Toma, Salvatore
; APPLICANT: Isacchi, Antonella
; TITLE OF INVENTION: Tankyrase Homolog Protein (THP), Nucleic Acids, and Methods Rel
; TITLE OF INVENTION: Same
; FILE REFERENCE: PHRM-0043
; CURRENT APPLICATION NUMBER: US/09/350,982C
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 1166
; TYPE: PRT
; ORGANISM: Artificial
```

INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 949 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-196-387-10

Query Match 8.9%; Score 209; DB 3; Length 949;
Best Local Similarity 27.4%; Pred. No. 9e-14;
Matches 93; Conservative 30; Mismatches 122; Indels 94; Gaps 11;
QY 139 AKNLMIDITKIFSLLOPDKEEDTTEKQALNQAVYDNDSTLDQLLRQERYKRFINSR 198
Db 628 AQMGNEAVQIILSESTPIR-TSDVD-----YRLLEASKAGDLETVKOLCSSQN-----VNCR 678
QY 199 SCMGVPGTPLRLAASVGHLSCLQVLLAHGADVSDLDVKAQTPLFTAVSHGHLDVVRVLE 258
Db 679 DLEGRHSTPLHFAAGYNNRVSVVEYLLHHGADVHAKDKGGLVPLHNACS YGHVEVAELLVR 738
QY 259 AGASPGGSYNNCSVPLTAARDGAVAILQELLDHGAENVK----- 299
Db 739 HGASVNVADLWKFTPLHEAAAKGYEICKLLKHGADPTKKNRDGNTPLDLVKEGDTDIQ 798
QY 300 -----AKLPVWA-----SNIASC-----SGELYLAAYVGHLDL 332
Db 799 DLLKGDAAALDAKKGCLARVQKLTPENI-NCRTQGRNSTPLHLAAGYNNLEVAEYLL 857
QY 333 LHGAD-----PDYNCCTDQ-GLLARVPRPTLLLEICLHNCPEYIQLLIDFGANI 381
Db 858 EHGADVNAQDKGGLIPLHNAASYGGCLARVQK-----LCTPENINCRDTQGRNS 906
QY 382 -----YLPSSLDLTSDQDKGIALLLQA 404
Db 907 TPLHLAAGYNNLEVAEYLLLEHGADVNAQDKGGLIPLHNA 945

RESULT 8

US-09-841-835-10
Sequence 10, Application US/09841835
Patent No. 6506587
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
TITLE OF INVENTION: OF USE THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/841,835
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/196,387
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-1684
TELEFAX: 201-343-1684

TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 949 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-841-835-10

Query Match 8.9%; Score 209; DB 4; Length 949;
Best Local Similarity 27.4%; Pred. No. 9e-14;
Matches 93; Conservative 30; Mismatches 122; Indels 94; Gaps 11;
QY 139 AKNLMIDITKIFSLLOPDKEEDTTEKQALNQAVYDNDSTLDQLLRQERYKRFINSR 198
Db 628 AQMGNEAVQIILSESTPIR-TSDVD-----YRLLEASKAGDLETVKOLCSSQN-----VNCR 678
QY 199 SCMGVPGTPLRLAASVGHLSCLQVLLAHGADVSDLDVKAQTPLFTAVSHGHLDVVRVLE 258
Db 679 DLEGRHSTPLHFAAGYNNRVSVVEYLLHHGADVHAKDKGGLVPLHNACS YGHVEVAELLVR 738
QY 259 AGASPGGSYNNCSVPLTAARDGAVAILQELLDHGAENVK----- 299
Db 739 HGASVNVADLWKFTPLHEAAAKGYEICKLLKHGADPTKKNRDGNTPLDLVKEGDTDIQ 798
QY 300 -----AKLPVWA-----SNIASC-----SGELYLAAYVGHLDL 332
Db 799 DLLKGDAAALDAKKGCLARVQKLTPENI-NCRTQGRNSTPLHLAAGYNNLEVAEYLL 857
QY 333 LHGAD-----PDYNCCTDQ-GLLARVPRPTLLLEICLHNCPEYIQLLIDFGANI 381
Db 858 EHGADVNAQDKGGLIPLHNAASYGGCLARVQK-----LCTPENINCRDTQGRNS 906
QY 382 -----YLPSSLDLTSDQDKGIALLLQA 404
Db 907 TPLHLAAGYNNLEVAEYLLLEHGADVNAQDKGGLIPLHNA 945

RESULT 9

US-09-196-387-2
Sequence 2, Application US/09196387
Patent No. 627613
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
TITLE OF INVENTION: OF USE THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/196,387
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/095,225
FILING DATE: June 10, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800


```

; ; TELFAX: 201-343-1684
; ; TELRX: 133521
; ; INFORMATION FOR SEQ ID NO:
; ; SEQUENCE CHARACTERISTICS
; ; LENGTH: 1327 amino acids
; ; TYPE: amino acid
; ; STRANDEDNESS: single
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: protein
; ; HYPOTHEetical: NO
; ; US-09-196-387-2

```

Query Match	8.9%;	Score	208.5;	DB	3;	Length	1327;
Best Local Similarity	26.7%;	Pred. No.	1.9e-13;				
Matches	100;	Conservative	41;	Mismatches	140;	Indels	93;
						Gaps	15;
139	AKGNLMDITKIFLLQPDKEEEDTDTBEKQALNQAVVDNSYTLDQLLRERYKRFNSR	198					
QY		DB					
628	AQNGNEAVQQLSESTPIR-TSDVD----	678	YRLLEASKAGDLTVKQLCSQN-----	VNCR			
QY		DB					
199	SGWGVGPTPURLAASYGHLSCLOVLLAHGADVPSLDVKAQTPPTFTAVSHGHLDVCVRVLE	258					
QY		DB					
679	DLSEGRHSTPLHFAAGYNRVSWVEYLLHHGADVHAKDKGGLVPLHNACSYGHYEVAELLR	738					
QY		DB					
259	AGASPGGSIYNNCSPVLTAARDGAVATLQELLDHGAENVK-----	299					
QY		DB					
739	HGASVNVADLWKVETPLHLEAAGAKYIECKULLXHGADFTKKNRDGNTPDLVKEGDDTDI	798					
QY		DB					
300	-----AKLPWA-----SNIASC-----SGPLYLAAYVCHLDCEFRLL	332					
QY		DB					
799	DLKGDAAALDAAKGCLARVQKLCITENI-NCRDTCGRNSTPLHLAAGYNNLEVAEYLL	857					
QY		DB					
333	LHGAD-----PDYNCCTDQG-----LLARVPRPRTLLEIC-----LHNCEPBYIQ	372					
QY		DB					
858	EHGADYNAQDKGGLIPLHNAASYGHVDIAALLIKYNTCVNATDKWAFTPLHEAAQKGRQT	917					
QY		DB					
373	----LLIDFGANIVLPSLS-----LDLTSODDKGIALLLQA-----RATPRSLLSOVRLLV	419					
QY		DB					
918	LCALLLAHGAD----PTWKNOEGQTPLDLATADDIR-ALLIDAMPEALPTCFKFPQATVVS	973					
QY		DB					
420	RRALCQAGQPOAIN	433					
QY		DB					
974	ASLISPASTPSCLS	987					
QY		DB					

RESULT 10
US-09-841-835-2
; Sequence 2, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387

[illegible]

```

RESULT 11
US-09-172-977-4
; Sequence 4, Application US/09172977
; Patent No. 5989863
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Guegier, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Yue, Henry
; TITLE OF INVENTION: HUMAN ANKYRIN FAMILY PROTEIN
; FILE REFERENCE: PF-0615 US
; CURRENT APPLICATION NUMBER: US/09/172,977
; CURRENT FILING DATE: 1998-10-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 1839
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: g29491

```

US-09-172-977-4

Query Match 8.6%; Score 202.5; DB 2; Length 1839;
Best Local Similarity 26.8%; Pred. No. 1.5e-12;
Matches 63; Conservative 31; Mismatches 84; Indels 57; Gaps 4;
QY 206 TPLRLAASVGHLSCLQVLLAHGADVDSLDVKAQTPLFTAVSHGHLDVCRVLLVLEAGASPGG 265
DB 565 TPLHVAAGYGLDVAKLLQRRRAADSAGKNGLTPLHVAHYDNQKVALLLLEKASPHA 624
QY 266 SIYNNCSPLVTAARDGAVAILQELLDHGAENVKAKLPVWASNIASCSGPLYLAAYVGH 325
DB 625 TAKNGYTPHLHIAAKNGQMOTASTLLNYGAETNIVTKQV-----TPLHLASQEGHT 675
QY 326 DCFRLLLHGAADPDYNTDQGLLARVPRPTLLEICLHNCEPEYIQLLIDFGANIYLP 385
DB 676 D-----MTLLDUGKGANIHMT 692
QY 386 ----LSLDTSDDK-GIALLQARATPRSLLSQVRLVVRALCOAGQPOAINQL 435
DB 693 KSLGTLHLAAQEDKVNADILTKHADQDAHTKLGVTPLIVACHYGNVMVNF 747

RESULT 12

US-09-082-059-2
; Sequence 2, Application US/09082059A
; Patent No. 6225086

; GENERAL INFORMATION:

; APPLICANT: Morrow, Jon S.

; APPLICANT: Devarajan, Prasad

; TITLE OF INVENTION: No. 6225086el Ankyrin Proteins and a Method for Their Identification

; FILE REFERENCE: 44574-5002-US

; CURRENT APPLICATION NUMBER: US/09/082,059A

; CURRENT FILING DATE: 1998-05-21

; EARLIER APPLICATION NUMBER: 60/047356

; EARLIER FILING DATE: 1997-05-21

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 1088

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-082-059-2

Query Match 8.6%; Score 202; DB 3; Length 1088;
Best Local Similarity 27.3%; Pred. No. 6.9e-13;
Matches 71; Conservative 34; Mismatches 89; Indels 66; Gaps 8;
QY 206 TPLRLAASVGHLSCLQVLLAHGADVDSLDVKA----- 237
DB 56 TPLHVAAGYGLDVAKLLQRRRAADSAGKNGLTPLHVAHYDNQKVALLLLEKASPHA 624
QY 238 ----OTPLFTAVSHGHLDVCRVLLVLEAGASPGSIYNNCSPLVTAARDGAVAILQELLDH 292
DB 116 KAKDDQTPHLHIAAKNGQMOTASTLLNYGAETNIVTKQV-----TPLHLASQEGHT 675
QY 293 GAEANYKAKLPVWASNIASCSG--PLYLAAYVGHLDVCRVLLVLEAGASPGSIYNNCSPLV 350
DB 176 GASLSITTK-----KGTTPHVAAGYGLDVAKLLQRRRAADSAGKNGLTPLHVAHYDNQK 220
QY 351 VPRPTLLEICLHNCEPEYIQLLIDFGA-----NIYLPSSLDTSDDKGIALL-- 401
DB 221 ----TPLHVAHYD-NQKVALLLDQASPHAAKNGYTPHLHIAAKNGQMOTASTLLLEY 274
QY 402 -LOARATPRSLLSQVRLVVR 420
DB 275 GADANAVTRQGIASVHLAAQ 294

RESULT 13

US-09-031-485-23

; Sequence 23, Application US/09031485

; Patent No. 5824306

; GENERAL INFORMATION:

; APPLICANT: Tang, Liang

; APPLICANT: Blehm, E. Scot

; TITLE OF INVENTION: DIOPHILARIA AND BRUGIA ANKYRIN

; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND

; TITLE OF INVENTION: USES THEREOF

; NUMBER OF SEQUENCES: 85

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Carol Talkington Verser, Ph.D.

; ADDRESSEE: Heska Corporation

; STREET: 1825 Sharp Point Drive

; CITY: Fort Collins

; STATE: Colorado

; COUNTRY: USA

; ZIP: 80525

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: Windows 95

; SOFTWARE: WordPerfect for Windows, Version 7.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/031,485

; FILING DATE:

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/847,429

; FILING DATE: 24-APR-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Verser, Carol Talkington

; REGISTRATION NUMBER: 37,459

; REFERENCE/DOCKET NUMBER: HW-5

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 970/493-7272

; TELEFAX: 970/484-9505

; INFORMATION FOR SEQ ID NO: 23:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 303 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-09-031-485-23

Query Match 8.6%; Score 201.5; DB 2; Length 303;
Best Local Similarity 26.9%; Pred. No. 9e-14;
Matches 66; Conservative 36; Mismatches 65; Indels 6;
QY 206 TPLRLAASVGHLSCLQVLLAHGADVDSLDVKAQTPLFTAVSHGHLDVCRVLLVLEAGASPGG 265
DB 109 TPLHVAAGYGLDVAKLLQRRRAADSAGKNGLTPLHVAHYDNQKVALLLLEKASPHA 168
QY 266 SIYNNCSPLVTAARDGAVAILQELLDHGAENVKAKLPVWASNIASCSG--PLYLAAYVGH 323
DB 169 ATRDLYTPHLHIAAKNGQBEVAAILMDHGTDKTLLTK-----KGTTPHLHIAAKY 217
QY 324 HLDCEFRLLHGAADPDYNTDQGLLARVPRPTLLEICLHNCEPEYIQLLIDFGANIY 383
DB 218 NLPVAKSLLETRTPVDIEGKQ-----VTPLHVAHYN----- 250
QY 384 PSLSLDLTSDDKGIALLQ-----ARATPRSLLSQVRLVVRALCOAGQPOAINQLDIPP 439
DB 251 ----NDKVALLLLENGASAHAAKNGYTPHLHIAAK-----NQMDIAS 289
QY 440 MLISY 444
DB 290 TLLHY 294

RESULT 14

US-08-847-429A-23

; Sequence 23, Application US/08847429A

; Patent No. 5827692

; GENERAL INFORMATION:

; APPLICANT: Tang, Liang

APPLICANT: Blehm, E. Scot
TITLE OF INVENTION: DIOFILARIA AND BRUGIA ANKYRIN
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
TITLE OF INVENTION: US\$ THEREOF
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/47,429A
FILING DATE: 24-APR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: HW-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-847-429A-23

Query Match 8.6%; Score 201.5; DB 2; Length 303;
Best Local Similarity 26.9%; Pred. No. 9e-14;
Matches 66; Conservative 36; Mismatches 78; Indels 65; Gaps 6;
QY 206 TPLRLAASYGHLSCLOVLLAHGADVSLDVKAQTPLFTAVSHGHLDVVRVLEAGASPGG 265
DB 109 TPLHLAARANQTDIVRVLRNGAQVDAARELQTPHLIASRLGNTDIVILLQANASPN 168
QY 266 SIYNNCSPLVTAARDGAVAILQELLDHGAENVKALPWASNIASCSG--PLYLAAYVG 323
DB 169 ATRDLYTPLHIAAKEGQEEVAAILMDHGTDKTLTK-----KGFTPLHIAAKYG 217
QY 324 HLDGCFRLLLHGAADPDYNTDQGLLARVPRPTLLEICLHNCPEPEYIQLLLDFGNIYL 383
DB 218 NLPVAKSLLERGTPVDIEGKNQ-----VTPHLVAAHYN----- 250
QY 384 PSLSLDLTSQDDKGIALLLQ-----ARATPRSLLSQVRLVRRALCQAGOPAINQLDIPP 439
DB 251 -----NDKVALLLENGASAHAAKNGYTPHLIAAK-----NQMDIAS 289
QY 440 MLISY 444
DB 290 TLLHY 294

RESULT 15
US-09-065-474-23
Sequence 23, Application US/09065474
Patent No. 6063599
GENERAL INFORMATION:
APPLICANT: Tang, Liang
APPLICANT: Blehm, E. Scot
TITLE OF INVENTION: DIOFILARIA AND BRUGIA ANKYRIN
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
TITLE OF INVENTION: US\$ THEREOF
NUMBER OF SEQUENCES: 171

CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/065,474
FILING DATE: 24-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: HW-5-C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-065-474-23

Query Match 8.6%; Score 201.5; DB 3; Length 303;
Best Local Similarity 26.9%; Pred. No. 9e-14;
Matches 66; Conservative 36; Mismatches 78; Indels 65; Gaps 6;
QY 206 TPLRLAASYGHLSCLOVLLAHGADVSLDVKAQTPLFTAVSHGHLDVVRVLEAGASPGG 265
DB 109 TPLHLAARANQTDIVRVLRNGAQVDAARELQTPHLIASRLGNTDIVILLQANASPN 168
QY 266 SIYNNCSPLVTAARDGAVAILQELLDHGAENVKALPWASNIASCSG--PLYLAAYVG 323
DB 169 ATRDLYTPLHIAAKEGQEEVAAILMDHGTDKTLTK-----KGFTPLHIAAKYG 217
QY 324 HLDGCFRLLLHGAADPDYNTDQGLLARVPRPTLLEICLHNCPEPEYIQLLLDFGNIYL 383
DB 218 NLPVAKSLLERGTPVDIEGKNQ-----VTPHLVAAHYN----- 250
QY 384 PSLSLDLTSQDDKGIALLLQ-----ARATPRSLLSQVRLVRRALCQAGOPAINQLDIPP 439
DB 251 -----NDKVALLLENGASAHAAKNGYTPHLIAAK-----NQMDIAS 289
QY 440 MLISY 444
DB 290 TLLHY 294

Search completed: December 1, 2003, 07:38:05
Job time : 44 secs

THIS PAGE BLANK (USPTO)

Db 347 GGGGCGAGGTGGACCTGGTGGATGTCAAGGGGACAGACTGCCCTGTATGTGGCTAGTG 406
QY 1105 CATGGGCATCTGACACTGTGTACGTGTCTTTTGGAGCTGGTCCCTCTCTCTGTGTAGC 1164
Db 407 AACGGGCACTTGGAGACACTGAGATCCTTTTGGAGCTGGTCTGATCCCAACGGCAGC 466
QY 1165 ATCTAACAACACTGTTCTCCCGTGTCTACAGCTGCCCGTGTATGTGTGTATCCTG 1224
Db 467 CGGCACACCGGACACTCTCTGTGTACCATGCTCTCTGTGTGGTAGGACGACATCCTG 526
QY 1225 CAGGAGCTCTAGACCATGTGTGAGGCGCAACGTCAAAGCTAAACT-----ACC 1274
Db 527 AAGGCTCTTATCAGGTATGGGAGATGTTGATGTCAACCATCATCTGAATCTGCACAC 586
QY 1275 AGTCTGGGATCAAAATAGTCTCATGTTCT-----GGCCCCCTCTATTGGCGCA 1326
Db 587 CGGCCCTCTTTTTCAGCGGCGCTAACTCTCTGGTGTCTGTCTCTATATCATCATGCT 646
QY 1327 GTCTACGGGCACTGGACTGTTTCCGCTCTCTTTCCTCACGGGGCAGACCCCTGACTAC 1386
Db 647 GCCTACCATCACTTTCAGTGTCTTTCAGGCTGCTCTTTCAGGCTGGGCAATCTGACTTC 706
QY 1387 AACTGCACTGACACGGC 1404
Db 707 AATTGCAATGGCCCTGTC 724

RESULT 2

US-09-302-769-26
; Sequence 26, Application US/09302769
; Patent No. 6323317
; GENERAL INFORMATION:
; APPLICANT: HILTON, Douglas J
; APPLICANT: ALEXANDER, Warren S
; APPLICANT: VINEY, Elizabeth M
; APPLICANT: WILLSON, Tracey A
; APPLICANT: RICHARDSON, Rachael T
; APPLICANT: STARR, Robyn
; APPLICANT: NICHOLSON, Sandra E
; APPLICANT: METCALF, Donald
; APPLICANT: NICOLA, Nicos A
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
; FILE REFERENCE: 109762
; CURRENT APPLICATION NUMBER: US/09/302,769
; CURRENT FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 08/962,560
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 419
; TYPE: DNA
; ORGANISM: Human
US-09-302-769-26

Query Match 3.8%; Score 71.2; DB 4; Length 419;
Best Local Similarity 55.4%; Pred. No. 2.2e-11;
Matches 164; Conservative 0; Mismatches 123; Indels 9; Gaps 1;
QY 862 GAGAAGCAGGCTCTCAATCAAGCAGTGTATGACACGACTCTATATCTTTGGACCAGCTT 921
Db 114 GAGGACACGAGGCTCATGATGAGCTTACGTCGGGACCTCAGACCCCTCAGGACCTTA 173
QY 922 TTGCGCCAGGAGCGTTTACAAACGTTTTCATCAACAGCAGGAGTGGCTGG-----GGT 972
Db 174 TTGCAAGAGGAGAGCTACCGAGCCGATCAACGAGAAGTCTGTCTGGTGTGTGGCTGG 233
QY 973 GTTCTGGGACACCTTGGCTTGGCTTCTTATATGGCCACTTGGCTGTGTGGAGTC 1032
Db 234 CTCCCTGCACACCGTTGGGAATCGCGGCCACTGCGAGGCCATGGGAGCTGTGTGGACTTC 293
QY 1033 CTCTTAGCCCATGGTGTGATGTGACAGCTTGGATGTCAAGGCACAGACGCCACTTTTC 1092

Db 294 CTCATCCGGAAGGGGGCCGAGGTGGATCTGTGGACGTAAAGACACAGACGCGCTGTAT 353
QY 1093 ACTGCTGTGAGTATGGCCATCTGGACTGTGTACGTGTGCTTTTGAAGCTGTGC 1148
Db 354 GTGGCTGTGTTGAACGGGCACCTAGAGAGTACCAGATCCTTCTCGAAGCTGCGGC 409
RESULT 3
US-09-221-298-28
; Sequence 28, Application US/09221298
; Patent No. 6284241
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471
; CURRENT APPLICATION NUMBER: US/09/221,298
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (212)
; OTHER INFORMATION: Where n is a, c, g or t
US-09-221-298-28

Query Match 3.7%; Score 69.4; DB 3; Length 401;
Best Local Similarity 52.4%; Pred. No. 7.4e-11;
Matches 176; Conservative 0; Mismatches 157; Indels 3; Gaps 1;
QY 360 AAAAGAGATGAGGAAAGTGGATGGAACCTGATTCACCAATATCAGACTTTGGGCGTAT 419
Db 52 AGAAAAGTTTAACTGGATGGATGGACAGTTCACAAATCCAGTGAAGAAATACAGGAGGCA 111
QY 420 GGAATACCCACAGAAACTGGACCATACAGATGCCAAGCAAGAACTATGAGATATGCAG 479
Db 112 GGGCTTGGCCCAATCACCATTGGAGAAATACCTTTTATTAAGTGTCTATGAGCTCTGCGA 171
QY 480 CACCTACCTCTGAAATAGTGTCTCTAAATCTGTCTACCTTTGGGAACGGTGTGGAAG 539
Db 172 CACTTACCTCTCTTTTGGTGTTCGGTATCGTGCCTCANATGATGACCTCCGAGAGT 231
QY 540 TTCAAAGTTTCAAGAAGTAAAGACGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 599
Db 232 TCAACTTTTAGTCCCGAAATCGAATTCAGTGTCTCATGATTCATCCAGAAATAA 291
QY 600 TGCTGCATTTGCCGTGTAGCCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 656
Db 292 GACGGTCAATTGTGCGTTCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 351
QY 657 TGATGAGCTCTTTTGGAGGCCATTAGCCAAACAAA 692
Db 352 TGATGAGAAATATCTCGATGTTATCAGGAGACTAA 387

RESULT 4

US-09-155-078-1
; Sequence 1, Application US/09155078
; Patent No. 6312688
; GENERAL INFORMATION:
; APPLICANT: POUSTKA, Annemarie
; APPLICANT: KIOSCHIS, Petra
; APPLICANT: LAPORTE, Jocelyn
; APPLICANT: HU, Ling Jia Louis
; APPLICANT: MANDEL, Jean Louis
; APPLICANT: DAHL, Niklas
; TITLE OF INVENTION: TYROSINE-PHOSPHATASE-RELATED PROTEIN
; FILE REFERENCE: 012627-010
; CURRENT APPLICATION NUMBER: US/09/155,078

;/ CURRENT FILING DATE: 1999-03-02
;/ EARLIER APPLICATION NUMBER: PCT/DE97/00592
;/ EARLIER FILING DATE: 1997-03-21
;/ EARLIER APPLICATION NUMBER: DE 196 11 234.6
;/ EARLIER FILING DATE: 1996-03-21
;/ NUMBER OF SEQ ID NOS: 4
;/ SOFTWARE: Patent in Ver. 2.0
;/ SEQ ID NO 1
;/ LENGTH: 3431
;/ TYPE: DNA
;/ ORGANISM: German-type microorganism & cell cul.
;/ FEATURE:
;/ NAME/KEY: CDS
;/ LOCATION: (1)..(1863)
US-09-155-078-1

Query Match 3.6%; Score 67.2; DB 4; Length 3431;
Best Local Similarity 52.1%; Pred. No. 1.3e-09;
Matches 175; Conservative 0; Mismatches 158; Indels 3; Gaps 1;
QY 360 AAAAGAGATGAGGGAAGTGGTGAACCTGATTGACCAATATACAGACTTTGGGGGTAT 419
DB 522 AGAAAGTTTACGTGGATGGATGGACAGTTTACATCCAGTGGAGAAATACAGAGGCA 581
QY 420 GGAATACCCACAGAAATCGACCAATACAGATGCCAACAGAAACTATGAGATATGCAG 479
DB 582 GGGCTTGCCCAATCACCATTGGAGAATAACTTTTATTATAAGTGTATGAGCTCTGTGA 641
QY 480 CACCTACCTCTGAAATAGTGGTTCCTAAATCTGTACTCTGGGACCGTGGTGGAG 539
DB 642 CACTTACCTCTCTTTTGGTGGTTCGGTATCGTGCCTCAGATGATGACCTCGGAGAGT 701
QY 540 TTCAAAGTTTCAAGTAAGAAAGCTGCTCCTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 599
DB 702 TCGAACTTTAGTCCCGAAATCGAATCCAGTGGTGTCTGATGATTCATCCAGAAATAA 761
QY 600 TCGTGCATTTGCCGTGTAGCCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 656
DB 762 GACGGTCAITTTGGTGTGCGTTCAGTCCAGCTCTTGTGCGTATGATGGGAAACGAAATAA 821
QY 657 TGATGAGCTCTTTGTTGGAGGCCATTAGCCAAACAA 692
DB 822 TGATGAGAAATATCTGATGTTTATCATCAGGAGACTAA 857

RESULT 5
US-08-181-629A-2
; Sequence 2, Application US/08181629A
; Patent No. 5472872
; GENERAL INFORMATION:
; APPLICANT: Swaminathan, Neela
; APPLICANT: Van Euten, James
; APPLICANT: Mead, David
; APPLICANT: Skowron, Piotr
; TITLE OF INVENTION: Recombinant CviJI Restriction Endonuclease
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/181,629A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:

;/ NAME: Clough, David W.
;/ REGISTRATION NUMBER: 36,107
;/ REFERENCE/DOCKET NUMBER: 31504
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: 312/474-6300
;/ TELEFAX: 312/474-0448
;/ TELEX: 25-3856
;/ INFORMATION FOR SEQ ID NO: 2:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 5496 base pairs
;/ TYPE: nucleic acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: DNA (genomic)
US-08-181-629A-2

Query Match 3.5%; Score 66.6; DB 1; Length 5496;
Best Local Similarity 52.7%; Pred. No. 2.5e-09;
Matches 144; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
QY 978 TGGGACACCCCTTGGCGCTTGGCTTCTTATGCGCACCTTGAGCTGTTTGCAGTCTCTT 1037
DB 3468 TTGGACGGGTTTACATTTACCGGCTTTTAAATGTCATAGCATGTCGTCAAGCGCTTAT 3527
QY 1038 AGCCCATGCTGCTGATGTTGACAGCTTGATGTCAGGCACAGACGCCACTTTTCACTGC 1097
DB 3528 TGATCGGGTGCAATCTTGACATCACAGATATTCGGGATGTACACCACTTCATCGTGC 3587
QY 1098 TGTGAGTCATGCCAATCTGGACTGTGTCGTGCTGCTTTTGGAGCTGGTCTCTCTGG 1157
DB 3588 GGTTTATAATGACACGATGTCATGTCGAAGATACCTCGTAGAAGCAGGTGCAACTCTTGA 3647
QY 1158 TGGTAGCATCTACACAACTGTTCTCCGCTGCTCACAGCTGCCGTGATGCTGCTTGC 1217
DB 3648 COTCATGATGATGTCAGTGGTGGCTTACATTACCGGGCTTTTAAATGGTAATGATGC 3707
QY 1218 TATCTCGAGGAGCTCTTAGACCATGTCAGCA 1250
DB 3708 GATTTTGAGGTGCTCATTTGAAGCAGGTGCAGA 3740

RESULT 6
US-09-509-802-1
; Sequence 1, Application US/09509802
; Patent No. 6489130
; GENERAL INFORMATION:
; APPLICANT: Immunex Corp.
; APPLICANT: Bird, Timothy
; APPLICANT: Virca, G.D.
; TITLE OF INVENTION: DEATH ASSOCIATED KINASE CONTAINING ANKYRIN REPEATS (DAKAR)
; FILE REFERENCE: 2889-US
; CURRENT APPLICATION NUMBER: US/09/509,802
; CURRENT FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 1
; LENGTH: 2370
; TYPE: DNA
; ORGANISM: Mus sp.
US-09-509-802-1

Query Match 3.2%; Score 60.2; DB 4; Length 2370;
Best Local Similarity 51.3%; Pred. No. 1.2e-07;
Matches 140; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
QY 980 GGACACCCCTTGGCGCTTGGCTTCTTATGCGCACCTTGAGCTGTTTGCAGTCTCTTAG 1039
DB 1829 GGACACCCCTTGCACCTGCTGCTCAGAGGGGCAATTACCGTGGCTCGCATTTCTATTG 1888
QY 1040 CCCATGGTCTGATGTTGACAGCTTGGATGTCAAGGCACAGACGCCACTTTTTCATCTGCTG 1099
DB 1889 ACCTGTGCTCTGATGTTAAATCTCGAGCTCTAGAGGCACAGACACCTCTGTCATGTTGCTG 1948

QY 1100 TCAGTCATGGCCATCTGGACTGTGTACGTGTCTTTGAAGCTGGTCTCTCCTGGTG 1159
|||
Db 1949 CAGAGACTGGACACACTAGTACTGCCAGCTACTCTTGCATCTGTGTGGCAAGGAGG 2008
|||
QY 1160 GTAGCATCTACAAACACTGTTCTCCCGTGTCTACAGCTGCCCGTGTGTGTCTTTGCTA 1219
|||
Db 2009 CTTTGACCTCAGAGGCTACTACTGCTTGCACCTGGCAGCCAGAGATGGACACCTTGGCTA 2068
|||
QY 1220 TCCTGAGAGGCTCTAGACCATGGTGCAGAGG 1252
|||
Db 2069 CTGTCAAGCTGCTCATAGAGGAGAGGCTGATG 2101
|||

RESULT 7

US-09-188-930-257
; Sequence 257, Application US/09188930A

; Patent No. 6150502

; GENERAL INFORMATION:

; APPLICANT: Watson, James D.

; APPLICANT: Strachan, Lorna

; APPLICANT: Sleeman, Matthew

; APPLICANT: Onrust, Rene

; APPLICANT: Murison, James Greg

; TITLE OF INVENTION: Compositions Isolated From Skin Cells

; FILE REFERENCE: 11000.1011c1

; CURRENT APPLICATION NUMBER: US/09/188,930A

; NUMBER OF SEQ ID NOS: 348

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 257

; LENGTH: 3516

; TYPE: DNA

; ORGANISM: Mouse

; US-09-188-930-257

Query Match 3.2%; Score 60.2; DB 3; Length 3516;

Best Local Similarity 51.3%; Pred. No. 1.6e-07;

Matches 140; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 980 GGACACCCCTTGGCTGGCTGCTTCTTATGGCCACTTGAGCTGTTTGAAGTCTCTTAG 1039
|||
Db 1823 GGACACCCCTGCACCTGGCTGCTCAGAGGGGGCATACCGTGGCTCGCATTCATTG 1882
|||

QY 1040 CCATGGTGTGATGTTGACAGCTTGGATGTCAAGGCACAGAGCCACTTTTCACTGCTG 1099
|||
Db 1883 ACCTGTGCTCTGATGTTAACTCTGCAGCCTACAGGCACAGACACCTCTGCATGTTGCTG 1942
|||

QY 1100 TCAGTCATGGCCATCTGGACTGTGTACGTGTCTTTTGAAGCTGGTCTCTCCTGGTG 1159
|||
Db 1943 CAGAGACTGGACACACTAGTACTGCCAGGCTACTCTTTCATCTGCTGGCAAGGAGG 2002
|||

QY 1160 GTAGCATCTACAAACACTGTTCTCCCGTGTCTACAGCTGCCCGTGTGTGTCTTTGCTA 1219
|||
Db 2003 CTTTGACCTCAGAGGCTACTACTGCTTGCACCTGGCAGCCAGAGATGGACACCTTGGCTA 2062
|||

QY 1220 TCCTGAGAGGCTCTAGACCATGGTGCAGAGG 1252
|||
Db 2063 CTGTCAAGCTGCTCATAGAGGAGAGGCTGATG 2095
|||

RESULT 8

US-09-312-283C-257

; Sequence 257, Application US/09312283C

; Patent No. 6573095

; GENERAL INFORMATION:

; APPLICANT: Watson, James D.

; APPLICANT: Strachan, Lorna

; APPLICANT: Sleeman, Matthew

; APPLICANT: Onrust, Rene

; APPLICANT: Murison, James G.

; APPLICANT: Kumble, Krishanand D.

; TITLE OF INVENTION: Compositions Isolated from Skin Cells

; TITLE OF INVENTION: and Methods for Their Use

; FILE REFERENCE: 11000.1011c2

; CURRENT APPLICATION NUMBER: US/09/312,283C

; CURRENT FILING DATE: 1999-05-14

; NUMBER OF SEQ ID NOS: 425

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 257

; LENGTH: 3516

; TYPE: DNA

; ORGANISM: Mouse

; US-09-312-283C-257

Query Match 3.2%; Score 60.2; DB 4; Length 3516;

Best Local Similarity 51.3%; Pred. No. 1.6e-07;

Matches 140; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 980 GGACACCCCTTGGCTGGCTGCTTCTTATGGCCACTTGAGCTGTTTGAAGTCTCTTAG 1039
|||
Db 1823 GGACACCCCTGCACCTGGCTGCTCAGAGGGGGCATACCGTGGCTCGCATTCATTG 1882
|||

QY 1040 CCATGGTGTCTGATGTTGACAGCTTGGATGTCAAGGCACAGAGCCACTTTTCACTGCTG 1099
|||
Db 1883 ACCTGTGCTCTGATGTTAACTCTGCAGCCTACAGGCACAGACACCTCTGCATGTTGCTG 1942
|||

QY 1100 TCAGTCATGGCCATCTGGACTGTGTACGTGTCTTTTGAAGCTGGTCTCTCCTGGTG 1159
|||
Db 1943 CAGAGACTGGACACACTAGTACTGCCAGGCTACTCTTTCATCTGCTGGCAAGGAGG 2002
|||

QY 1160 GTAGCATCTACAAACACTGTTCTCCCGTGTCTCAGAGCTGCCCGTGTGTGTGTTGCTA 1219
|||
Db 2003 CTTTGACCTCAGAGGCTACTACTGCTTGCACCTGGCAGCCAGAGATGGACACCTGGCTA 2062
|||

QY 1220 TCCTGAGAGGCTCTAGACCATGGTGCAGAGG 1252
|||
Db 2063 CTGTCAAGCTGCTCATAGAGGAGGCTGATG 2095
|||

RESULT 9

US-09-392-812A-3

; Sequence 3, Application US/09392812A

; Patent No. 6537778

; GENERAL INFORMATION:

; APPLICANT: Zuker, Charles S.

; APPLICANT: Walker, Richard G.

; APPLICANT: Willingham, Aaron

; TITLE OF INVENTION: The Regents of the University of California

; FILE REFERENCE: 023078-097600US

; CURRENT APPLICATION NUMBER: US/09/392,812A

; CURRENT FILING DATE: 1999-09-09

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 3

; LENGTH: 6156

; TYPE: DNA

; ORGANISM: Drosophila melanogaster

; FEATURE:

; OTHER INFORMATION: nompC cDNA sequence

; US-09-392-812A-3

Query Match 2.9%; Score 55.4; DB 4; Length 6156;

Best Local Similarity 49.8%; Pred. No. 6.1e-06;

Matches 140; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 980 GGACACCCCTTGGCTGGCTGCTTCTTATGGCCACTTGAGCTGTTTGAAGTCTCTTAG 1039
|||
Db 3755 GGACGGGCTGCATATCGCGCCATCATCGATGGCCACATCCAGATGGTGGAGATTCGCTCG 3814
|||

QY 1040 CCATGGTGTCTGATGTTGACAGCTTGGATGTCAAGGCACAGAGCCACTTTTCACTGCTG 1099
|||
Db 3815 GCCAGGGCGGAGATCAACCGAACCAGATCGGAACGGTTTGAAGCTGGTCTCTCCTGGTG 3874
|||

QY 1100 TCAGTCATGGCCATCTGGACTGTGTACGTGTCTTTTGAAGCTGGTCTCTCCTGGTG 1159
|||

RESULT 12

US-09-196-387-1
; Sequence 1, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4134 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; US-09-196-387-1

Query Match 2.7%; Score 51.4; DB 3; Length 4134;
Best Local Similarity 49.1%; Pred. No. 7.5e-05;
Matches 136; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 983 CACCCTTGCGCTTGGCTGCTTCTTATGGCCACTTGAGCTGTTTGGCAAGTCCTCTTAGCCC 1042
DB 658 CTCCTCGCACTTCGCTGCAGGTTTGGGAAGGAAGGATGTTGTAGAACACTTACTACAGA 717
QY 1043 ATGCTGCTGATGTTGACAGCTTGATGTCAGGCAAGCCACCTTTTCACCTGCTGTCA 1102
DB 718 TGGGTGCTAATGTCACGCTCGTGATGATGAGGCTCTCATCCCGCTTCAATGCTGTT 777
QY 1103 GTCATGGCCATCGGACTGTGTACGTGCTGCTTTTGGAAAGCTGCTCTCTCGTGGTGA 1162
DB 778 CTTTGGCCATGCTGAGTTGTGAGTCTGTTATTTGGCCAGGAGCTGATCCAAATGCCA 837
QY 1163 GCATCTACAACTGTTCTCCCGTGTCTACAGCTGCCCGTGTGCTGCTGTTGCTATCC 1222
DB 838 GGGATAACTGGAACCTATACACCTCTGCATGAAGCTGCTATTAAAGGAAGATCGATGTGT 897
QY 1223 TGCAGAGCTCTAGACCATGTTGTCAGAGGCCAAAGT 1259
DB 898 GCATTGTGCTGTCAGACGAGCTGACCCAAACAT 934

RESULT 13

US-09-841-835-1
; Sequence 1, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4134 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; US-09-841-835-1

Query Match 2.7%; Score 51.4; DB 4; Length 4134;
Best Local Similarity 49.1%; Pred. No. 7.5e-05;
Matches 136; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 983 CACCCTTGCGCTTGGCTGCTTCTTATGGCCACTTGAGCTGTTTGGCAAGTCCTCTTAGCCC 1042
DB 658 CTCCTCGCACTTCGCTGCAGGTTTGGGAAGGAAGGATGTTGTAGAACACTTACTACAGA 717
QY 1043 ATGCTGCTGATGTTGACAGCTTGGATGTCAAGCCACAGCCACCTTTTCACCTGCTGTCA 1102
DB 718 TGGGTGCTAATGTCACGCTCGTGATGATGAGGCTCTCATCCCGCTTCAATGCTGTT 777
QY 1103 GTCATGGCCATCGGACTGTGTACGTGCTGCTTTTGGAAAGCTGCTCTCTCGTGGTGA 1162
DB 778 CTTTGGCCATGCTGAGTTGTGAGTCTGTTATTTGGCCAGGAGCTGATCCAAATGCCA 837
QY 1163 GCATCTACAACTGTTCTCCCGTGTCTACAGCTGCCCGTGTGCTGCTGTTGCTATCC 1222
DB 838 GGGATAACTGGAACCTATACACCTCTGCATGAAGCTGCTATTAAAGGAAGATCGATGTGT 897
QY 1223 TGCAGAGCTCTAGACCATGTTGTCAGAGGCCAAAGT 1259
DB 898 GCATTGTGCTGTCAGACGAGCTGACCCAAACAT 934

RESULT 14

US-09-196-387-7

; Sequence 7, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4491 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6..2027
; US-09-196-387-7

Query Match 2.7%; Score 51.4; DB 3; Length 4491;
Best Local Similarity 49.1%; Pred. No. 7.9e-05;
Matches 136; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 983 CACCCCTGGCTGGCTGCTTCTTATGGCCACTTGAGCTGTTTGGCAAGTCCTCTTAGCCCC 1042
Db 658 CTCCTCGTGCATCTCGCTGAGGTTTGGGAAGGAGGATGTTGAGAACACTTACTACAGA 717
QY 1043 ATGGTGCTGATGTTGACAGCTTGATGGAGGCTCTCATCCGCTTCTATAATGCTGTT 777
Db 718 TGGGTGCTAAATGCTCCACCGCTCGTGATGGAGGCTCATCCGCTTCTATAATGCTGTT 777
QY 1103 GTCATGGCCATCTGAGCAGTGTGTGCTGCTTTTGGGAAGCTGGTGCCTCTCTGGTGGTA 1162
Db 778 CTTTGGCCATCTGAGTGTGTGAGTCTGTTATTTGGCAAGAGCTGATCCAAATGCCA 837
QY 1163 GCATCTACAACTGTTCTCCCGTGCTCACAGCTGCCGCTGATGCTGTTGCTATCC 1222
Db 838 GGGATAACTGGAACATATACACCTCTGCATGAAGCTCTATTAAAGGAAGATCGATGCT 897
QY 1223 TGCAGAGCTCCTAGACCATGTTGCGAGGCCAACGT 1259
Db 898 GCATTGTGCTGTCAGCAGCGAGCTGACCCAAACAT 934

RESULT 15

US-09-841-835-7
; Sequence 7, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4491 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6..2027
; US-09-841-835-7

Query Match 2.7%; Score 51.4; DB 4; Length 4491;
Best Local Similarity 49.1%; Pred. No. 7.9e-05;
Matches 136; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 983 CACCCCTGGCTGGCTGCTTCTTATGGCCACTTGAGCTGTTTGGCAAGTCCTCTTAGCCCC 1042
Db 658 CTCCTCGTGCATCTCGCTGAGGTTTGGGAAGGAGGATGTTGAGAACACTTACTACAGA 717
QY 1043 ATGGTGCTGATGTTGACAGCTTGATGGAGGCTCTCATCCGCTTCTATAATGCTGTT 1102
Db 718 TGGGTGCTAAATGCTCCACCGCTCGTGATGGAGGCTCATCCGCTTCTATAATGCTGTT 777
QY 1103 GTCATGGCCATCTGAGCAGTGTGTGCTGCTTTTGGGAAGCTGGTGCCTCTCTGGTGGTA 1162
Db 778 CTTTGGCCATCTGAGTGTGTGAGTCTGTTATTTGGCAAGAGCTGATCCAAATGCCA 837
QY 1163 GCATCTACAACTGTTCTCCCGTGCTCACAGCTGCCGCTGATGCTGTTGCTATCC 1222
Db 838 GGGATAACTGGAACATATACACCTCTGCATGAAGCTCTATTAAAGGAAGATCGATGTT 897
QY 1223 TGCAGAGCTCCTAGACCATGTTGCGAGGCCAACGT 1259

Db 898 GCATTGTGCTGCTGCAGCAGGAGCTGACCCAAACAT 934

Search completed: December 1, 2003, 13:33:36
Job time : 141 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 2, 2003, 08:30:46 ; Search time 0.001 Seconds
(without alignments)
1686.444 Million cell updates/sec

Title: us-09-941-831a-20
Perfect score: 2343
Sequence: 1 MRESGWKLDIPDSFGRMG1.....QAINQLDIPPLISYLKHL 449

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1 seqs, 1878 residues

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 1 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=soft -Q=us-09-941-831a-20 -DB=us-09-941-831a-6
-SUFFIX=pro -OUT=align20_6 -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0 -UNITS=bits
-START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=1 -DOCALIGN=200
-THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=1 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -NCPU=6 -NO_XLPXY
-NEG SCORES=0 -LONGLOG -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXIT=7

Database : us-09-941-831a-6:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2343	100.0	1878	1	us-09-941-831a-6

ALIGNMENTS

RESULT 1
us-09-941-831a-6

Alignment Scores:

Pred. No.:	0	Length:	1878
Score:	2343.00	Matches:	449
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

us-09-941-831a-20 (1-449) x us-09-941-831a-6 (1-1878)

Qy 1 MetArgLusGlyTrpLysLeuLeuAspProileSerAspPheGlyArgMetGlyIle 20
Db 367 ATGAGGGAAAGTGGATGGAAACCTGATTGACCAATATCAGACTTTGGGCGTATGGGAATA 426

Qy	21	ProAsnArgAsnTrpThrIleThrAspAlaAsnArgAsnTyrGluIleCysSerThrTyr	40
Db	427	CCCAACAGAACTGGACCATACACATGCCAACAAACATATGAGATATGCAGACCTAC	486
Qy	41	ProProGluIleValProLysSerValThrLeuGlyThrValValGlySerSerLys	60
Db	487	CCTCTGAAATAGTGGTTCCTAAATCTGTACCTTGGGAACGGTGGTTGGAAGTTCAAA	546
Qy	61	PheArgSerLysGluArgValProValLeuSerTyrLeuTyrLysGluAsnAlaAla	80
Db	547	TTCAAGAGTAAAGAACGGTCTCTGTCTCTCTACCTCTCAAAAGAGAACAAATGCTGCC	606
Qy	81	IleCysArgCysSerGlnProLeuSerGlyPheTyrThrArgCysValAspAspGluLeu	100
Db	607	ATTTCGCGCTGTAGCCAGCCCTCTCTGTGATTTTACACTCGCTGTGTAGATGATGAGCTC	666
Qy	101	LeuLeuGluAlaIleSerGlnThrAsnProGlySerGlnPheMetTyrValValAspThr	120
Db	667	TTGTTGGAGGCCATTTAGCCAAACAAACCCAGGAGCCAGTTTATGTATGTTGTAGACACA	726
Qy	121	ArgProLysIleTrpHisPheLeuValLeuIleMetArgIleValLeuGlnLeuAlaLys	140
Db	727	AGACCAAGATCTGGCATTTCTGTGTCTCAATAGAGATAGTTCTCAATTAGCCAAAG	786
Qy	141	MetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGlu	160
Db	787	ATGAACCTCATGGACATCACCAAGATCTTCTCCCTCTGCAGCCGACAGGAGGAG	846
Qy	161	AspThrAspThrGluGluGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyr	180
Db	847	GACACTGACACAGAGAGAGCAGGCTCTCAATCAAGCAGTGTATGACAAACGACTCTAT	906
Qy	181	ThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGly	200
Db	907	ACTTTGGACCAAGCTTTTGGCCAGGAGCGTTACAAACGTTTCATCAACAGCAGGAGTGC	966
Qy	201	TrpGlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeu	220
Db	967	TGGGTGTCTCTGGGACACCTTGGCTGGCTGCTTCTATGGCCACTTGGAGCTGTTTG	1026
Qy	221	GlnValLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrPro	240
Db	1027	CAAGTCTCTTAGCCCATGGTGTGATGTGACAGCTTGGATGTCAAGGCACAGACGCCA	1086
Qy	241	LeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGly	260
Db	1087	CTTTTCACTGCTGTGAGTCTGGCCATCTGGACTGTGACGTGTGCTTTTGGAAAGCTGT	1146
Qy	261	AlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAsp	280
Db	1147	GCCTCTCTCTGGTGTGAGTATCTACAAACACTGTTCTCCGCTGCTCACAGCTCCCGTAT	1206
Qy	281	GlyAlaValAlaIleLeuGlnLeuLeuAspHisGlyAlaGluAlaAsnValLysAla	300
Db	1207	GGTGTGTTGCTATCTCTGCAGGAGCTCTTAGACCATGGTGCAGAGGCCAACGCTCAAAGCT	1266
Qy	301	LysLeuProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAla	320
Db	1267	AAACTACCAGTCTGGGCATCAAAACATAGCTTCTGTGGCCCTCTCTATTGTCGCCA	1326
Qy	321	ValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyValAspProAspTyr	340
Db	1327	GTCACGGGACCTCGAGCTGTTTCCGCTGCTTTTGTCTCCACGGGCGACAGCCCTGACTAC	1386
Qy	341	AsnCysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIle	360
Db	1387	AACTGCACTGACAGGGCTATTGGCTCGTGCTCCCAAGACCCCGCACCCCTCTTGAATC	1446
Qy	361	CysLeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsn	380
Db	1447	TGCCTCCATCATATTTGTGAGCCAGATATATCCAGCTGTTTAATCGATTTTGTGTGTAAT	1506

```
Qy 381 I l e T y r L e u P r o S e r L e u S e r L e u A s p L e u T h r S e r G l n A s p L y s G l y I l e A l a L e u 400
    |||||
Db 1507 A T C T A C C T T C C A T C T C T C C C T T G A C C T G A C C T C A C A G A T G A T A A A G G C A T T G C A T T G 1566

Qy 401 L e u L e u G l n A l a A r g A l a T h r P r o A r g S e r L e u S e r G l n V a l A r g L e u V a l V a l A r g 420
    |||||
Db 1567 C T G C T A C A G G C C G A G C C A C T C C A C G G T C A C T T C T A T C A C A G G T C C G T T T A G T C G T C C G C 1626

Qy 421 A r g A l a L e u C y s G l n A l a G l y G l n P r o G l n A l a I l e A s n G l n L e u A s p I l e P r o P r o M e t 440
    |||||
Db 1627 A G A G C C T T G T G C C A G G C T G C C C A G C C A A G C C A T C A A C C A G C T G G A T A T T C T C C C A T G 1686

Qy 441 L e u I l e S e r T y r L e u L y s H i s G l n L e u 449
    |||||
Db 1687 T T G A T T A G C T A C C T A A A A C A C C A A C T G 1713
```

Search completed: December 2, 2003, 08:30:48
Job time : 2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 1, 2003, 09:49:33 ; Search time 3611 Seconds
(without alignments)

12640.212 Million cell updates/sec

Title: US-09-941-831A-6

Perfect score: 1878

Sequence: 1 catgattacccaagcttgg.....taaaaaaaaaaaaaaaaaaaaaa 1878

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:**

1: em_estba:**

2: em_esthum:**

3: em_estin:**

4: em_estnu:**

5: em_estov:**

6: em_estpl:**

7: em_estro:**

8: em_htc:**

9: gb_est1:**

10: gb_est2:**

11: gb_htc:**

12: gb_est3:**

13: gb_est4:**

14: gb_est5:**

15: em_estfun:**

16: em_estom:**

17: em_gss_hum:**

18: em_gss_inv:**

19: em_gss_pln:**

20: em_gss_vrt:**

21: em_gss_fun:**

22: em_gss_man:**

23: em_gss_mus:**

24: em_gss_pro:**

25: em_gss_rod:**

26: em_gss_phg:**

27: em_gss_vrl:**

28: gb_gss1:**

29: gb_gss2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	729.2	38.8	1262	11 AK009644	AK009644 Mus muscu
2	675.2	36.0	1201	9 AL529411	AL529411 AL529411
3	644	34.3	644	13 BX107270	BX107270 BX107270
4	617	32.9	625	10 AW957331	AW957331 EST369521

5	593.8	31.6	719	12 BM190670	BM190670 POSM01000
6	532.2	28.3	943	10 BF037514	BF037514 601460634
7	498.8	26.6	678	29 AG127205	AG127205 Pan trogl
8	482	25.7	482	9 AA633395	AA633395 AC33304.8
9	452.2	24.1	543	14 W72063	W72063 ZDB66f07.81
10	445	23.7	446	14 W76256	W76256 ZDB66f07.81
11	404.6	21.5	464	14 W58499	W58499 ZD20C05.81
12	388.8	20.7	579	12 BI344684	BI344684 373212 MA
13	368.6	19.6	859	9 AJ454588	AJ454588 AJ454588
14	368.4	19.6	518	28 AZ725513	AZ725513 RPCI-24-1
15	358	19.1	362	9 AA349955	AA349955 EST56937
16	344.2	18.3	932	14 BY709303	BY709303 BY709303
17	339	18.1	342	9 AA349954	AA349954 EST56936
18	332.2	17.7	751	13 BU383527	BU383527 603859981
19	331	17.6	643	10 BB661935	BB661935 BB661935
20	330.2	17.6	722	9 AJ451253	AJ451253 AJ451253
21	329	17.5	627	10 BB662275	BB662275 BB662275
22	314.8	16.8	659	12 BM439955	BM439955 pgrin.pk0
23	310.8	16.5	908	13 BU220575	BU220575 603106716
24	297.6	15.8	524	14 W58534	W58534 ZD20C05.81
25	297.4	15.8	861	13 BU254916	BU254916 603747645
26	293.8	15.6	538	12 BI290522	BI290522 UI-R-DK0-
27	279.4	14.9	372	28 AZ725227	AZ725227 RPCI-24-1
28	275	14.7	849	13 BU228932	BU228932 603798270
29	275	14.6	713	14 CA364256	CA364256 639132 NC
30	269.8	14.4	486	10 BF889720	BF889720 289285 MA
31	269.4	14.3	758	13 BX073379	BX073379 BX073379
32	265.8	14.2	755	14 CA368575	CA368575 644799 NC
33	265.4	14.1	680	14 CA358977	CA358977 631831 NC
34	258.8	13.8	990	12 BM545864	BM545864 AGENCOURT
35	249.2	13.3	318	9 AA778891	AA778891 ZJ46503.8
36	241.4	12.9	1060	29 CC217390	CC217390 CH261-79C
37	233.6	12.4	2165	11 AK081973	AK081973 Mus muscu
38	232.6	12.4	798	13 BQ179015	BQ179015 UI-M-EVO-
39	229.6	12.2	567	10 BF949440	BF949440 MR3-NN021
40	228	12.1	3814	11 AK076218	AK076218 Mus muscu
41	226.4	12.1	837	13 BQ946838	BQ946838 AGENCOURT
42	226	12.0	690	13 BU382590	BU382590 603858336
43	224.2	11.9	2471	11 AK028160	AK028160 Mus muscu
44	223.8	11.9	436	12 BI290849	BI290849 UI-R-DK0-
45	223.8	11.9	464	10 BF555515	BF555515 UI-R-A1-d

ALIGNMENTS

RESULT 1	AK009644	1262 bp	mRNA	linear	HTC 05-DEC-2002
LOCUS	Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310036C05				
DEFINITION	SOCs box protein 12, full insert sequence.				
ACCESSION	AK009644				
VERSION	AK009644.1	GI:12844562			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	1				
AUTHORS	Carninci, P., and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2				
AUTHORS	Carninci, P., Shibata, Y., Hayateu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				


```
Db 625 TTGCCATCTTACAGGAGCTCTAGGCGATGGTCCGAGGCTAATGTCAAAGCTAAACTAC 684
Qy 1274 CAGTCTGGGATCAAAACATAGCTTCATGTTCTGGCCCTCTATTTGGCGGAGTCTACG 1333
Db 685 CAGTCTGGGCGTCAAAATATAGCTTCATGTTCTGGCCCTCTATTTGGCGGAGTCTATG 744
Qy 1334 GGCACCTGGAAGTCTTTCCGCTCTTTTGTCTCCACGGGGCAGACCTGACTACAACTGCA 1393
Db 745 GGCACCTTGATGTTTCCGCTCTTTTGTCTATGGGCGAGATCTGATTACAACTGCA 804
Qy 1394 CTGACAGGGCCTATTGGCTGTGTCCTCCAAAGACCCGCGACCTCTCTGAAATCTGCTCC 1453
Db 805 CTGACAGGGCCTTTTAAAGTCGTGTTCCACAGGCTCGCACACTCTCTGAAATCTGCTTC 864
Qy 1454 ATCATATTTGTGAGCCAGAGTATATCCAGCTGTATATCGATTTTGGTGTATATCTACC 1513
Db 865 ATCATATTTGTGAGCCAGAGTATATCCAGCTGTATATCGATTTTGGTGTATATCTACC 924
Qy 1514 TTCCATCTCTCTCCCTTGACCTGACCTCAAGATGATAAGGCAATGCAATGCTCTAC 1573
Db 925 TTCCATCTCTCTCTGTCGACCACTTCAAGATGATAAGGCAATGCAATGCTCTAC 984
Qy 1574 AGCCCGAGCACTCCAGGTCATCTTATCAAGTCCGTTTGTAGTCCGCGAGACCT 1633
Db 985 AGCCCGAGCACTCCAGGTCATCTCTGTCGACGACCCGTTTGTAGTATCCGCGAGATCC 1044
Qy 1634 TGTGCGAGCTGCGGACGACACAGGATCAAGGTCGATGATGATGATGATGATGATGATGAT 1693
Db 1045 TGTGCGGCGGACCAAGTCAAGGTCGATGATGATGATGATGATGATGATGATGATGAT 1104
Qy 1694 GCTACCTAAACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1748
Db 1105 GCTACCTAAACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1164
Qy 1749 CGAAAAA-CCACCTGGGAGTCACTGATGATGATGATGATGATGATGATGATGATGATGAT 1807
Db 1165 CAAAAAATAAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1224
Qy 1808 GGAG 1811
Db 1225 GGCG 1228

RESULT 2
AL529411 1201 bp mRNA linear EST 23-MAY-2003
LOCUS AL529411 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens
DEFINITION cDNA clone CS0DD006Y19 5-PRIME, mRNA sequence.
ACCESSION AL529411
VERSION AL529411.2 GI:31067254
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 13, 2001 this sequence version replaced gi:12792904.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
10696. r For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DD006AG10QPI&cluster=10696.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DD006AG10QPI.
```

```
FEATURES
source
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/no_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DD006Y19"
/tissue_type="NEUROBLASTOMA COT 50-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"
/Note="First strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 309 a 241 c 276 g 326 t 49 others
ORIGIN
Query Match 36.0%; Score 675.2; DB 9; Length 1201;
Best Local Similarity 95.3%; Pred. No. 1.2e-96;
Matches 682; Conservative 20; Mismatches 13; Indels 1; Gaps 1;
Qy 23 CGAGGTAGAAAACGTAATTTGGTGGATCGTTATGTAGTAAAGAAACAGCTAATGGGA 82
Db 136 CCAAGGTAGAAAACGTAATTTGGTGGATCGTTATGTAGTAAAGAAACAGCTAATGGGA 195
Qy 83 TTCTTTATCTTACTGCAACCCACCTGATCTATGTGGAGGCTTCAGGTGCAGCCCGAAAG 142
Db 196 TTCTTTATCTTACTGCAACCCACCTGATCTATGTGGAGGCTTCAGGTGCAGCCCGAAAG 255
Qy 143 AAACATGATGATGCACTCCATCATTCACATGTCACATGTCGAGAGTATCCCATCAGCTGG 202
Db 256 WWACATGATGATGCACTCCATCATTCACATGTCGAGAGTATCCCATCAGCTGG 315
Qy 203 GTTGTCCCTGACCTCCGCTGCAAGAAATTTCCGGTGGCCACCTTTGTTTATGATTCTG 262
Db 316 GTTGTCCCTGACCTCCGCTGCAAGAAATTTCCGGTGGCCACCTTTGTTTATGATTCTG 375
Qy 263 ACCTTGTGTGCACTGAGGTTTATTTTCACTGCTCAAGCTTTCTCAGCCAGCATTTACCTG 322
Db 376 ACCTTGTGTGCACTGAGGTTTATTTTCACTGCTCAAGCTTTCTCAGCCAGCTTTACCTG 435
Qy 323 AAGATCTTTATGCTTTTCTTATATCCCAATCTCCAAAGAGATGAGGGAAGTGGAT 382
Db 436 AAGATCTTTATGCTTTTCTTATATCCCAATCTCCAAAGAGATGAGGGAAGTGGAT 495
Qy 383 GGAAACTGATGACCCCAATATCAGACTTTGGGCGTATGGGAATCCCAACAGATCTGGA 442
Db 496 GGTAACTGATGACCCCAATATCAGACTTTGGGCGTATGGGAATCCCAACAGATCTGGA 555
Qy 443 CCATAACAGATGCCAACAGAAATCTAGATATGACAGACCTTACCTCTCTGAAATAGTG- 501
Db 556 CCATAACAGATGCCAACAGAAATCTAGATATGACAGATGCTCTCTCTCTGATATAGTCT 615
Qy 502 GTTCTTAATCTGTTTACCTTGGGAAACGGTGGTGAAGTTCAAAGTTCAAGAGTAAAGAA 561
Db 616 GTTCTTAATCTGTTTACCTTGGGAAACGGTGGTGAAGTTCAAAGTTCAAGAGTAAAGAA 675
Qy 562 CGTGTCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 621
Db 676 CGTGTCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 735
Qy 622 CAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 681
Db 736 CAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 795
Qy 682 AGCCAAACAAACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 737
Db 796 AGCCAAACAAACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 851
RESULT 3
BX107270 644 bp mRNA linear EST 06-FEB-2003
LOCUS BX107270 Soares fetal heart Nbh19W Homo sapiens cDNA clone
DEFINITION IMAGP998N09778 ; IMAGE:341192, mRNA sequence.
```



```
QY 78 TGGGATCTTTATCTTACTGCAACCACTGATCTATGTGGAGGCTTCAGGTGCGAGCCG 137
Db 61 TGGGATCTTTATCTTACTGCAACCACTGATCTATGTGGAGGCTTCAGGTGCGAGCCG 120
QY 138 GAAAGAAACATGATGATGCACTCATCATCATCATCATCATCATCATCATCATCATCAT 197
Db 121 GAAAGAAACATGATGATGCACTCATCATCATCATCATCATCATCATCATCATCATCAT 180
QY 198 CTGGGTTGTCCTCCCTGACCTCGCTGCAAGAAATTTCCGGGTGGCCCACTTTGTTTGA 257
Db 181 CTGGGTTGTCCTCCCTGACCTCGCTGCAAGAAATTTCCGGGTGGCCCACTTTGTTTGA 240
QY 258 TTCTGACCTTGTGTGTCATGAGGTTTATATTTTCACTGTCTCAAGCTTTCTCAGCCAGCAT 317
Db 241 TTCTGACCTTGTGTGTCATGAGGTTTATATTTTCACTGTCTCAAGCTTTCTCAGCCAGCAT 300
QY 318 ACCTGAAGATCTTTATGCTTTTCTTTATATATCCAAATCTCTCAAAAGAGATGAGGGAAG 377
Db 301 ACCTGAAGATCTTTATGCTTTTCTTTATATATCCAAATCTCTCAAAAGAGATGAGGGAAG 360
QY 378 TGGATGGAACATGATGACCAATATCAGACTTTGGGCGTGTGGGAATACCCAAACAGAA 437
Db 361 TGGATGGAACATGATGACCAATATCAGACTTTGGGCGTGTGGGAATACCCAAACAGAA 420
QY 438 CTGGACCAATACAGATGACCAACAGAACTATGAGATATGAGATATGAGATATGAGATATGAGAT 497
Db 421 CTGGACCAATACAGATGACCAACAGAACTATGAGATATGAGATATGAGATATGAGATATGAGAT 480
QY 498 AGTGGTTCTTAATCTGTTACCTTGGGAACGGTGGTGGAGTTCAAAGTTCAAAGTTAA 557
Db 481 AGTGGTTCTTAATCTGTTACCTTGGGAACGGTGGTGGAGTTCAAAGTTCAAAGTTAA 540
QY 558 AGAACCTGTCCTGCTCTCTACCTCTACAAAGAGAACTGCTGCCATTTGCCGCTG 617
Db 541 AGAACCTGTCCTGCTCTCTACCTCTACAAAGAGAACTGCTGCCATTTGCCGCTG 600
QY 618 TAGCCAGCCTCTCTGGAATTTTAC 642
Db 601 GAGCCAGCCTTCTGGAATTTTAC 625
```

```
RESULT 5
BM190670 719 bp mRNA linear EST 11-DRC-2001
LOCUS POSM0100011.E08F porcine skeletal muscle cDNA library (POSM) Sus
DEFINITION scrofa cDNA 5', mRNA sequence.
ACCESSION BM190670
VERSION BM190670.1 GI:17526633
KEYWORDS EST.
SOURCE Sus scrofa (pig)
```

```
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 719)
Yao J., Coussens, P., Ernst, C. and Saama, P.
AUTHORS Analysis of expressed sequence tags from a normalized porcine
TITLE skeletal muscle cDNA library
```

```
JOURNAL Unpublished
COMMENT Contact: Jianbo Yao
Department of Animal Science
Michigan State University
B215 Anthony Hall, East Lansing, MI, USA
Tel: 517-353-8443
Fax: 517-353-1699
Email: yaoj@msu.edu
Seq primer: M13 reverse.
```

```
FEATURES
source 1..719
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/sex="male and female"
/tissue_type="skeletal muscles"
```

```
/dev_stage="45 d and 90 d of gestation, birth, 7 wk and 1
yr of age"
/lab_host="DH10B"
/clone_lib="porcine skeletal muscle cDNA library (POSM)"
/notes="Organ: hind limbs; Vector: pSPORT1; Site_1: NotI;
Site_2: SalI"
BASE COUNT 151 a 206 c 176 g 182 t 4 others
ORIGIN
```

```
Query Match 31.6%; Score 593.8; DB 12; Length 719;
Best Local Similarity 90.0%; Pred. No. 9.1e-84;
Matches 645; Conservative 0; Mismatches 71; Indels 1; Gaps 1;

QY 818 CCTCTCTGAGCCGACCAAGGAGGAGGAGGACCTGACACAGAGAGAGAGGCTCTCA 877
Db 2 CAGGGGTCCGGCCCAAGGAGGAGGAGGACCTGACACAGGAGAGAGGAGGAGGCTCA 61
QY 878 ATCAAGCAGTGTATGACAAACGACTCCTATACCTTTGGACGAGCTTTTGGCCAGGAGCGTT 937
Db 62 ATCAAGCAGTGTATGACAAACGACTCCTATACCTTTGGACGAGCTTTTGGCCAGGAGCGAT 121
QY 938 ACAAGCCTTTCATCAACAGCAGGAGTGGC-TGGGGGTGTTCCCTGGGACACCCCTTGGCTTG 996
Db 122 ACAAGCAGTTCATCAATAGTAGGAGTGGCTTGGGGGTGTCCTGGGACACCCCTTGGCTTG 181
QY 997 GCTGCTTCTTATGGCCACTTGGAGCTGTTTGCAGTCTCTTAGCCCATGGTGTCTGATGTT 1056
Db 182 GCTGCTTCTTATGGCCACTTANCTTGTCTTGGAGTCTCTTGGCAGCATGGTGTCTGAGCT 241
QY 1057 GACAGCTTGGATGTCAAGGACAGAGCGACATTTTTCATGCTGTGTAGTCAATGGCCATCTG 1116
Db 242 GACAGCTTGGAGTCAAGGACAGAGCACCATTTTATCCGCGCTGAGTCAATGGCCATCTG 301
QY 1117 GACTGTGACGTGTCTTTTGGAAAGCTGGTGTCTCTCTGGTGGTAGCATCTACACAAAC 1176
Db 302 GACTGTGCGTGTCTTTTGGAAAGCTGGTGTCTCTCTGGTGGTAGCATCTACACAAAC 361
QY 1177 TGTTCTCCCGTCTCACAGCTGCCGTGATGTGTCTGTGTGTCTCTCTGAGGAGCTCTTA 1236
Db 362 TGCTCTCTCTGCTCACAGCTGCCGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 421
QY 1237 GACCATGTGTGAGAGGCGCAACGTCAAAGCTAAACTACAGTCTGGGCATCAAAACATAGCT 1296
Db 422 GGTATGTGTGAGAGGCGCAATGTCAAGGCAAAACTACCAAGTCTGGTCTATCAAAACATAGCT 481
QY 1297 TCATGTTCTGGCCCTCTCTATTTGGCCGAGTCTACGGGACCTGGAGCTGTTTCCGCTG 1356
Db 482 TCATGTTCTGGCCCTCTCTATTTGGCTGCAGTCTATGTGTGCTGCTGCTGCTGCTGCTGCT 541
QY 1357 CTTTGTCTCCAGGGGAGAGCCCTGACTACAACTGCACTGACAGGCGCTTATGGCTCGT 1416
Db 542 CTTTGTCTCCAGGAGAGATCTGACTATACTGACTATACTGACTGATCAACACCTGTGGCTCGA 601
QY 1417 GTCCCAAGACCCGCGACCCCTCTCTTGAATCTGCCTCCATCATATAATGTGAGCCAGAGTAT 1476
Db 602 GTCCCAAGACCCGCGACCCCTCTTAAATTTGCCCTCCACATAAATGTGAGCCAGAGTAT 661
QY 1477 ATCCAGCTGTTAATCGATTTTGGTGTGTAATATCTACCTTCCATCTCTCTCCCTTGAC 1533
Db 662 ATCCAGCTGTTAATGATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 718
```

```
RESULT 6
BF037514
LOCUS 601460634F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3864128 5',
DEFINITION mRNA sequence.
ACCESSION BF037514
VERSION BF037514.1 GI:10745763
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```

REFERENCE
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE        1 (bases 1 to 943)
JOURNAL      NIH-MGC http://mhc.nci.nih.gov/.
COMMENT      National Institutes of Health, Mammalian Gene Collection (MGC)
              Unpublished
              Contact: Robert Strausberg, Ph.D.
              Email: cgabbs@mail.nih.gov
              cDNA Library Preparation: Life Technologies, Inc.
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: L1AM9605 row: c column: 09
              High quality sequence stop: 579.

FEATURES
source
1. .943
   /organism="Homo sapiens"
   /mol_type="mRNA"
   /db_xref="taxon:9606"
   /clone="IMAGE:3864128"
   /tissue_type="adenocarcinoma"
   /lab_host="DH10B (phage-resistant)"
   /clone_lib="NIH_MGC_66"
   /notes="Organ: ovary; Vector: pCMV-SPORT6; Site 1: NotI;
   Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
   Average insert size 1.8 kb. Library constructed by Life
   Technologies."
BASE COUNT 254 a 217 c 244 g 227 t 1 others
ORIGIN
Query Match      28.3%; Score 532.2; DB 10; Length 943;
Best Local Similarity 89.5%; Pred. No. 3.7e-74;
Matches 620; Conservative 0; Mismatches 63; Indels 10; Gaps 4;

QY 23 CGAGGGTAGAAAACGTGAATGGTGATCGTTATGTGAGTAAGAAACCAAGCTTAATGGGA 82
DB 67 CCAAGGTAGAAAACGTGAATGGTGATCGTTATGTGAGTAAGAAACCAAGCTTAATGGGA 126
QY 83 TTCTTTATCTTACTGCAACCCACTGATCTATGTGAGGCTTCAGGTGCGAGCCCGGAAG 142
DB 127 TTCTTTATCTTACTGCAACCCACTGATCTATGTGAGGCTTCAGGTGCGAGCCCGGAAG 186
QY 143 AAACATGGATTGCACTCCATCATCTGCACTGCGAGAGTTACCCATCATCTAGCTGG 202
DB 187 AAACATGGATTGCACTCCATCATCTGCGAGAGTTACCCATCATCTAGCTGG 246
QY 203 GTTGTCCCTGACCTCCGCTGCAAGAAATTCGCGGTGGCCCACTTTGTTTAGATTCTG 262
DB 247 GTTGTCCCTGACCTCCGCTGCAAGAAATTCGCGGTGGCCCACTTTGTTTAGATTCTG 306
QY 263 ACCTGTGTCATGAGGTTTATTTACTGCTCAAGCTTCTCAGCCAGCATATTACCTG 322
DB 307 ACCTGTGTCATGAGGTTTATTTACTGCTCAAGCTTCTCAGCCAGCATATTACCTG 366
QY 323 AAGATCTTTATGCTTTTCTTAATCCCAATCTCAAAAGAGATGAGGAAAGTGAT 382
DB 367 AAGATCTTTATGCTTTTCTTAATCCCAATCTCAAAAGAGATGAGGAAAGTGAT 426
QY 383 GGAACATGATTGACCAATATCAGACTTTGGCGTATGGGAATPACCAACAGAAATCGGA 442
DB 427 GGAACATGATTGACCAATATCAGACTTTGGCGTATGGGAATPACCAACAGAAATCGGA 486
QY 443 CCATACAGATGCAACAGAACTATGAGATATGAGACCTTACCCTCTGAAATAGTGG 502
DB 487 CCATACAGATGCAACAGAACTATGAGATATGAGACCTTACCCTCTGAAATAGTGG 546
QY 503 TTCCTAAATCTGTTTACCTTGGGAAACCGTGGTTGGAAGTTCAAAAGTTCAAGAGTAAAGAAC 562
DB 547 TTCCTAAATCTGTTTACCTTGGGAAACCGTGGTTGGAAGTTCAAAAGTAAAGAAC 605
QY 563 GTGTCCTCTGTCTCCTACCTCTACAAAGAGAAACAAATGCTGCCATTTGCCCGGTGAGCC 622

```

```

606 GTGTCC---TGTGGTCTCTAACTCTACAAGAGAACAAATGTCG---ATTGGGGGTGAGCC 660
623 AGCTCTCTCTGATTTTACACTCGCTGTGTAGATGATGAGCTCTTGTGTGGAGGCCATTA 682
661 AGCTCTCTCGGATTTTACCTCGCTGGTGAATTAAGAGCCT---GGGGGAGGCACTA 716
683 GCCAAACAAACCCAGGAGCGAGTTCATGATG 715
717 GACAAAACACAGGGGGCGGTGTTGTGTG 749

RESULT 7
AG127205/c
LOCUS      678 bp      DNA      linear      GSS 04-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-137P23.F, genomic survey sequence.
ACCESSION  AG127205
VERSION     AG127205.1 GI:16656370
KEYWORDS   GSS.
SOURCE     Pan troglodytes (chimpanzee)
ORGANISM   Pan troglodytes
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE  1
AUTHORS    Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE      BAC end sequences of Library PTB
JOURNAL    Unpublished
AUTHORS    2 (bases 1 to 678)
            Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE      Direct Submission
JOURNAL    Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail: chimps@isc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT    Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector      : pKS145
R.Site 1    : SacI
R.Site 2    : SacI.
Location/Qualifiers
1. .678
   /organism="Pan troglodytes"
   /mol_type="genomic DNA"
   /db_xref="taxon:9598"
   /clone="PTB-137P23.F"
   /sex="male"
   /cell_type="lymphoblast"
   /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 174 a 182 c 178 g 143 t 1 others
ORIGIN
Query Match      26.6%; Score 498.8; DB 29; Length 678;
Best Local Similarity 97.7%; Pred. No. 7.7e-69;
Matches 506; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 730 CCAAGATCTGCATTCCTTGTGCTCATTAATGAGATAGTTCTCCAATTAGCCAGATG 789
DB 586 CCACAGATCTGCATTCCTTGTGCTCATTAATGAGATAGTTCTCCAATTAGCCAGATG 527
QY 790 AACCTCATGGACATCACCAGATCTTCTCCCTCTCGAGCCGCAAGAGGAGGAGGAC 849
DB 526 AACCTCATGGACATCACCAGATCTTCTCCCTCTCGAGCCGCAAGAGGAGGAGGAC 467
QY 850 ACTGACACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 909
DB 466 ACTGACACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 407

```

```

QY 910 TTGGACCACTTTTGGCCAGGAGCGTTTCAAAACGTTTTCATCAACAGCAGGAGTGCGTGG 969
|
|
|
Db 406 TTGGACCACTTTTGGCCAGGAGCGTTTCAAAACGTTTTCATCAACAGCAGGAGTGCGTGG 347
|
|
|
QY 970 GGTGTTCTTGGGACACCCCTTGGCTTGGCTTGGCTTCTTATGGCCACTTGGAGCTGTTTGC 1029
|
|
|
Db 346 GCGGTTTCTTGGGACACCCCTTGGCTTGGCTTGGCTTCTTATGAACACTTGGAGCTGTTTGC 287
|
|
|
QY 1030 GTCCCTTTAGCCCATCGTCTGATGTTGACAGCTTGTGATGTTCAAGSCACAGACGCCACTT 1089
|
|
|
Db 286 GTCCCTTTAGCCCATCGTCTGATGTTGACAGCTTGTGATGTTCAAGSCACAGACGCCACTT 227
|
|
|
QY 1090 TTCACTGCTGTGAGTCAATGCCATCTGAGCTGTGTACGTTGTGTTTGAAGCTGTGTGCC 1149
|
|
|
Db 226 TTCACTGCTGTGAGTCAATGCCATCTGAGCTGTGTACGTTGTGTTTGAAGCTGTGTGCC 167
|
|
|
QY 1150 TCTCTGTGTGTAGTCAATCTAACAACACTGTTTCTCCGCTGTCAAGCTGCCCGTGTATGTT 1209
|
|
|
Db 166 TGTCTGTGTGTAGTCAATCTAACAACACTGTTTCTCCGCTGTCAAGCTGCCCGTGTATGTT 107
|
|
|
QY 1210 GCTGTTGTCTATCTCGCAGGAGCTCTAGACCATGTTGTC 1247
|
|
|
Db 106 GCTGTTGTCTATCTCGCAGGAGCTCTCCGCAACGTCGTAC 69
|
|
|
RESULT 8
AA633985/c
LOCUS
DEFINITION
ac33f04.s1 Stratagene hnt neuron (#937233) Homo sapiens cDNA clone
IMAGE:858271 3', mRNA sequence.
ACCESSION
AA633985
VERSION
AA633985.1 GI:2557199
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 482)
AUTHORS
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wyllie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
TITLE
Unpublished
JOURNAL
Contact: Wilson RK
COMMENT
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LML ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1829 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 446.
FEATURES
Location/Qualifiers
1..482
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:858271"
/dev_stage="hnt neurons"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene hnt neuron (#937233)"
/notes="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; Cloned unidirectionally. Primer: Oligo dT.
Differentiated, post mitotic hnt neurons. Average insert
size: 1.5 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5'
GTAATCGGCAGCAG 3' ~3' adaptor sequence: 3'
CTCAGATTTTTTTTTTTTTTTT 3'"
BASE COUNT 121 a 88 c 153 g 120 t
ORIGIN

```

```

Query Match 25.7%; Score 482; DB 9; Length 482;
Best Local Similarity 100.0%; Pred. No. 3.9e-66;
Matches 482; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1374 AGACCTGACTACAACTGACCTGACGAGGCGCTATTGGCTGGTGTCCCAAGACCCCGCAC 1433
|
|
|
Db 482 AGACCTGACTACAACTGACCTGACGAGGCGCTATTGGCTGGTGTCCCAAGACCCCGCAC 423
|
|
|
QY 1434 CTTCTCTGAAATCTGCTCCATCAATAATTGTGAGCCAGAGTATATCCAGCTGTTAATCGA 1493
|
|
|
Db 422 CTTCTCTGAAATCTGCTCCATCAATAATTGTGAGCCAGAGTATATCCAGCTGTTAATCGA 363
|
|
|
QY 1494 TTTTGGTGTCTAATATCTACCTTCCATCTCTCTCCCTTGACCTGACCTCACAAGATGATAA 1553
|
|
|
Db 362 TTTTGGTGTCTAATATCTACCTTCCATCTCTCTCCCTTGACCTGACCTCACAAGATGATAA 303
|
|
|
QY 1554 AGGCATTGCTGCTGCTACAGGCCCGAGCCACTCCAGGTCACCTTCTATCAAGTCCG 1613
|
|
|
Db 302 AGGCATTGCTGCTGCTACAGGCCCGAGCCACTCCAGGTCACCTTCTATCAAGTCCG 243
|
|
|
QY 1614 TTTAGTCTGCTGCGCAGAGCTTGTGCGAGCTGGCCAGCCACCAAGCATCAACAGCTGA 1673
|
|
|
Db 242 TTTAGTCTGCTGCGCAGAGCTTGTGCGAGCTGGCCAGCCACCAAGCATCAACAGCTGA 183
|
|
|
QY 1674 TATTCCTCCATGTTGATTAGCTACCTAAACACCAACTGTAATCTTGAGTCTCCCCAG 1733
|
|
|
Db 182 TATTCCTCCATGTTGATTAGCTACCTAAACACCAACTGTAATCTTGAGTCTCCCCAG 123
|
|
|
QY 1734 GAACATTATGATGCTCCGAAACACCTGGGAGCTCAGTAGCTGGAGAGCATTTACAGCC 1793
|
|
|
Db 122 GAACATTATGATGCTCCGAAACACCTGGGAGCTCAGTAGCTGGAGAGCATTTACAGCC 63
|
|
|
QY 1794 TCATCCACTTACCTGGAGCTGCTCTCTCTATTATCTCCCAATAAATTTCTCCAGAA 1853
|
|
|
Db 62 TCATCCACTTACCTGGAGCTGCTCTCTCTATTATCTCCCAATAAATTTCTCCAGAA 3
|
|
|
QY 1854 AT 1855
|
|
|
Db 2 AT 1

RESULT 9
W72063/c
LOCUS
DEFINITION
zd66f07.s1 Soares fetal_heart_NbHH19W Homo sapiens cDNA clone
IMAGE:345637 3', mRNA sequence.
ACCESSION
W72063
VERSION
W72063.1 GI:1382333
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 543)
AUTHORS
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevisakis,E., Waterston
R., Williamson,A., Wohldmann,P. and Wilson,R.
The WashU-Merck EST Project
TITLE
Unpublished
JOURNAL
Contact: Wilson RK
COMMENT
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LML ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 945 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 360.
FEATURES
Location/Qualifiers

```


QY 1213 GTTGCTATCTCGCAGGAGCTCTAGACCATGGTGAGAGCCCAACGCTAAAGCTAAACTA 1272
 Db 242 GTTGCTATCTCGCAGGAGCTCTAGACCATGGTGAGAGCCCAACGCTAAAGCTAAACTA 301
 QY 1273 CCAGTCTGGGCATCAACATAGCTTTCATGTTCTGGCCCCCTCTATTGTCGCCGAGTCTAC 1332
 Db 302 CCAGTCTGGGCATCAACATAGCTTTCATGTTCTGGCCCCCTCTATTGTCGCCGAGTCTAC 361
 QY 1333 GGGCACTTGGACTGTTTTCGCCCTGCTTTTGGCTCCAGGGGAGACCCCTGACTACAACTGC 1392
 Db 362 GGGCACTTGGACTGTTTTCGCCCTGCTTTTGGCTCCAGGGGAGACCCCTGACTACAACTGC 421
 QY 1393 ACTGACCGGGCTATTGGCTCGTG 1417
 Db 422 ACTGACCGGGCTATTGGCTCGTG 446

RESULT 11
 W58499/c
 LOCUS
 DEFINITION IMAGE:341192 3', mRNA sequence.
 ACCESSION W58499.1 GI:1365358
 VERSION EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 464)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
 M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
 Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston
 R., Williamson, A., Wohlmann, P. and Wilson, R.
 The WashU-Merck EST Project
 Unpublished
 Contact: Wilton RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LLM; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 979 Std Error: 0.00
 Seq primer: mob.RSGA-ET
 High quality sequence stop: 351.
 Location/Qualifiers
 1. 464
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="GBB:1266567"
 /db_xref="taxon:9606"
 /clone="IMAGE:341192"
 /sex="unknown"
 /dev_stage="19 weeks"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares_fetal_heart_NDH19W"
 /notes="Organ: heart; Vector: pT73D (Pharmacia) with a
 modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTTACCAATCTGAAGTGGAGCGCGCATCTTTTCTTTTCTTTTCTTTT 3']
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by
 M.Fatima Bonaldo. This library was constructed from the
 same fetus as the fetal lung library, Soares fetal lung
 NDH19W."

FEATURES
 source

BASE COUNT
 ORIGIN

Query Match 21.5%; Score 404.6; DB 14; Length 464;
 Best Local Similarity 96.7%; Pred. No. 5.7e-54;
 Matches 444; Conservative 0; Mismatches 11; Indels 4; Gaps 3;
 QY 1401 GGGCCATATGGCTCGTGTGCCAAGAGCCCGGACCCCTCTTGAATCTGCCTCCATCATATA 1460
 Db 459 GGCNTTATGGGCTCGTGTGCCAAGAGCCCGGACCCCTCTTGAATCTGCCTCCATCATATA 400
 QY 1461 TTGTGAGCCA-CAGTATATCCAGCTG--TTAATCGATTTTGGTGTGAATATCTACCTTCC 1517
 Db 399 TTGTGAGCCAGAGTATATCCAGCTGGTTAATCGGATTTTGGTGTGAATATCTACCTTCC 340
 QY 1518 ATCTCTCTCCCTTGGACCTGACCTCAAGATGATAAAGCATTTGCTGTCTACAGGC 1577
 Db 339 ATCTCTCTCCCTTGGACCTGACCTCAAGATGATAAAGCATTTGCTGTCTACAGGC 280
 QY 1578 CCGAGCCACTCCAGGTCACTTCTATCATCAGGTCCGTTTAGTCTCCGCGAGAGCCTTGTG 1637
 Db 279 CCGAGCCACTCCAGGTCACTTCTATCATCAGGTCCGTTTAGTCTCCGCGAGAGCCTTGTG 220
 QY 1638 CC-AGGCTGGCCAGCCACAAGCCATCAACACAGCTGGATTTCTCCCATGTTGATTAGCT 1696
 Db 219 CCNAGGCTGGCCAGCCACAAGCCATCAACACAGCTGGATTTCTCCCATGTTGATTAGCT 160
 QY 1697 ACCTAAACACCAACTGTAATCTTGCAGTCTCCCGAGGAACCTTATGATGCCCTCCGAAAC 1756
 Db 159 ACCTAAACACCAACTGTAATCTTGCAGTCTCCCGAGGAACCTTATGATGCCCTCCGAAAC 100
 QY 1757 CACCTGGGACTCAGTGTAGCTGGAGAGCATTTACAGCTCATCCATCTTACCTGGAGCTGCT 1816
 Db 99 CACCTGGGACTCAGTGTAGCTGGAGAGCATTTACAGCTCATCCATCTTACCTGGAGCTGCT 40
 QY 1817 CTCCTGTATTATCTCCCAATAAATAATCTCCAGAAAAT 1855
 Db 39 CTCCTGTATTATCTCCCAATAAATAATCTCCAGAAAAT 1
 RESULT 12
 BI344684 579 bp mRNA linear EST 30-JUL-2001
 LOCUS 373212 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
 DEFINITION BI344684
 ACCESSION BI344684
 VERSION BI344684.1 GI:15037973
 KEYWORDS EST.
 SOURCE Sus scrofa (pig)
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 Vallet, J., Wise, T., Smith, T.P.L., Freking, B.A., Cho, J., White, J.,
 Fahrenkrug, S.C., Smith, T.P.L., Rohrer, G.A., Perle, G., Sultana, R., Quackenbush
 J., and Keele, J.W.
 Porcine gene discovery by normalized cDNA-library sequencing and
 EST cluster assembly
 Mamm. Genome 13 (8), 475-478 (2002)
 22213789
 12226715
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt_trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -minmatch 12 options.
 PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCAGTCAGCAGC
 Plate: 119 row: K column: 23
 Seq primer: ATTAGGTGACACTATAG.
 Location/Qualifiers
 1. 579
 source

```
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon.9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="WARC 2P1G"
/notes="vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT 155 a 127 c 153 g 144 t
ORIGIN
Query Match 20.7%; Score 388.8; DB 12; Length 579;
Best Local Similarity 91.7%; Pred. No. 1.5e-51;
Matches 411; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
QY 23 CGAGGGTAGAAAACGTTGGTGAATCGTTATGTGAGTAGAAGAACCCAGCTAATGGGA 82
Db 132 CCAAGGTAGAAAACGTTGAATTAATCTGGATCGTTGTGTGGGTAAAGAAACCCAGCTAATGGGA 191
QY 83 TTCTTTATCTTACTGCAACCCACCTGATCTATGTGGAGGCTTCAGGTGAGCCCGGAAG 142
Db 192 GTCTATACCTTATGCAACCCACCTGATCTATGTGGAGGCTTCGTGTGCTGCGCCGAAAG 251
QY 143 AAACATGGATTGCACTCCATCCATGCACTGCGAGAGTTTACCCATCAGCTAGCCTGG 202
Db 252 AGACATGGATTGCACTGCATCATATATGCGCACTGTGAGAGTTGCTTATTAATCTACCTGG 311
QY 203 GTTGTCCCTGACCTCCGCTGCAAGAAATTCGGGTGGCCCACTTTGTTTATAGATTCTG 262
Db 312 GTTGTCCCTGATCCTCCACTGCAAGAAATTCGGGTGGCCCACTTTGTTTATAGATTCTG 371
QY 263 ACCTTGTGCGCATGAGTTTATATTTCACTGCTCAAGCTTTCTCAGCCAGCATTAACCTG 322
Db 372 ACCTTGTGCGCATGAGTTTATATTTCACTGCTCAAGCTTTCTCAGCCAGCATTAACCTG 431
QY 323 AAGATCTTTATGCTTTTCTTATAATCCCAATCTCAAAAGAGATGAGGGAAGTGGAT 382
Db 432 AAGATCTTTATGCTTTTCTTATAATCCCAATCTCAAAAGAAATGAGGGAAGTGGAT 491
QY 383 GGAACCTGATTGACCAATATCAGACTTTGGGGGTATGGGAATACCCACAGAAATCTGA 442
Db 492 GGAGACTGATTGACCAATATCAGACTTTGGAGCTATGGGAATACCTAACAGGTACTGG 551
QY 443 CCATACAGATGCCACAGAACTATGA 470
Db 552 CCATACAGATGCCACAGAACTATGA 579
```

```
RESULT 13
AJ454588
LOCUS AJ454588 riken1 Gallus gallus cDNA clone 3d12r1, mRNA sequence.
DEFINITION AJ454588
ACCESSION AJ454588.1 GI:20264684
VERSION AJ454588.1
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 859)
Buerstedde,J.M.
Gallus gallus bursal lymphocyte EST
Unpublished
Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
Location/Qualifiers
1..859
/organism="Gallus gallus"
```

```
/mol_type="mRNA"
/db_xref="taxon:9031"
/clone="3d12r1"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/clone_lib="riken1"
/notes="CB inbred strain"
BASE COUNT 234 a 196 c 226 g 198 t 5 others
ORIGIN
Query Match 19.6%; Score 368.6; DB 9; Length 859;
Best Local Similarity 70.5%; Pred. No. 1.9e-48;
Matches 491; Conservative 0; Mismatches 205; Indels 0; Gaps 0;
QY 23 CGAGGGTAGAAAACGTTGGTGAATCGTTATGTGAGTAGAAGAACCCAGCTAATGGGA 82
Db 110 CGAAGGTGAAAACGTTGAATTAATCTAGATCGCTTATCAAAATAGGAAGGAGCAAGCGGGA 169
QY 83 TTCTTTATCTTACTGCAACCCACCTGATCTATGTGGAGGCTTCAGGTGAGCCCGGAAG 142
Db 170 CGTTATACCTGACGCCACCCATCTCATCTATGTGGATGCTTCTGCTGAAGTCAGAAAG 229
QY 143 AAACATGGATTGCACTCCATCAATGCACTGTGGAGAAAGTTACCCATCAGCTAGCTGG 202
Db 230 AAACATGGATTGCACTCACCATAATTGCACTGTAGAAAAATTCCTCTGACCCACAGCTG 289
QY 203 GTTGTCCCTGACCTCCGCTGCAAGAAATTCGGGTGGCCCACTTTGTTTATAGATTCTG 262
Db 290 GATGCCCACTCCTTATATTCACGCAAGAACTTCCACGTGGCTCACTTTGTTTATGGGCAGG 349
QY 263 ACCTTGTGCGCATGAGTTTATATTTCACTGCTCAAGCTTTCTCAGCCAGCATTAACCTG 322
Db 350 AAGCTGATTGTCAGAAAGTGTATACCTCTGCTCAAACTTTCACACAGGTGAACCCG 409
QY 323 AAGATCTTTATGCTTTTCTTATAATCCCAATCTCAAAAGAGATGAGGGAAGTGGAT 382
Db 410 AAGAACTTTATGCTTTTCTTAAATCTTAAAGATGCTTAAAGATAACCGGAGATTGGAT 469
QY 383 GGAACCTGATTGACCCATATCAGACTTTGGGGGTATGGGAATACCCACAGAACTGA 442
Db 470 GGAAGCTGATTGATTGAAAGTAGATTACCAAGCTATGGGAATCCCGAATGACTATTGGG 529
QY 443 CCATACAGATGCCACAGAACTATGAGATATGAGACACTTACGACACTTACCCTCTCGAAATAGTGG 502
Db 530 AATTAACAGATCTTAAACAAAGACTACGAGTTTGGCAACACATATCTCCAGAAATCGTGG 589
QY 503 TTCTTAAATCTGTTTACCTTTGGAAACGGTGGTGGAAAGTTCAAAGTTCAAGAGTAAAGAAC 562
Db 590 TGCCTCGAGCTGCTAGTAAAGCAACAGTATTGGAAGCTCAAGGTTTCAAGACGAGGGGGC 649
QY 563 GTGTCCTGTGCTCTTACCTTCAAAAGAAACAATGCTGCCATTTCCTGCTAGCC 622
Db 650 GGATCCCAAGTGTCTTCTTACTTATATAAGAAAAACAATGCTGCCATATGCTGCTGTAGCC 709
QY 623 AGCCTCTCTCTGGATTTTACACTCGCTGTGTAGATGATGAGCTCTTGTGGAGGCCATTGA 682
Db 710 AGCCCTCTCTGGGTTTCANTGCTCGTTGTTGGAGATGAACAATGTTGACAGGCATCA 769
QY 683 GCCAAACAAACCCAGGAGCCAGTTTATGTATGTTG 718
Db 770 GAGAAGCCAAACCTGGGAGTCCCTTCATGTTGTTG 805
```

```
RESULT 14
AZ725513
LOCUS AZ725513
DEFINITION AZ725513
ACCESSION AZ725513
VERSION AZ725513.1 GI:12472241
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
```

AZ725513 518 bp DNA linear GSS 24-JAN-2001
RPCI-24-145H23.TJ RPCI-24 Mus musculus genomic clone RPCI-24-145H23
genomic survey sequence.

Qy	268	GTGTGCATGAGTTTATATTTCACTGCTCAAGCTTTCTCAGCCAGCATTACCTGAAGAT	327
Db	241	GTGTGCATGAGTTTATATTTCACTGCTCAAGCTTTCTCAGCCAGCATTACCTGAAGAT	300
Qy	328	CTTTATGCTTTTCTTATATCCCAATCCTCAAAAGAGATGAGGGAAGTGGATGGAAA	387
Db	301	CTTTATGCTTTTCTTATATCCCAATCCTCAAAAGAGATGAGGGAAGTGGATGGAAA	360
Qy	388	CT 389	
Db	361	CT 362	

Search completed: December 1, 2003, 13:31:09
Job time : 3620 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2003, 10:38:24 ; Search time 70 Seconds
(without alignments)
1655.223 Million cell updates/sec

Title: US-09-941-831A-20
Perfect score: 2343
Sequence: 1 MRESGWKLDIPISDFGRMGI.....QAINQLDIPPMISYLKQL 449

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	DB ID	Description
1	1636	69.8	318 4 Q8N8F5	Q8N8F5 homo sapien
2	1435	61.2	308 11 Q9D738	Q9D738 mus musculus
3	675.5	28.8	704 4 Q9GEFO	Q9GEFO homo sapien
4	673	28.7	401 4 Q9Y3N3	Q9Y3N3 homo sapien
5	480.5	20.5	327 11 Q9D9S9	Q9D9S9 mus musculus
6	421	18.0	617 11 Q8VE11	Q8VE11 mus musculus
7	408	17.4	621 4 Q9EP80	Q9EP80 homo sapien
8	396	16.9	467 11 Q8K0A2	Q8K0A2 mus musculus
9	396	16.9	660 11 Q8C4J6	Q8C4J6 mus musculus
10	360	15.4	676 5 Q9W1Q6	Q9W1Q6 drosophila
11	360	15.4	761 5 Q8MLR7	Q8MLR7 drosophila
12	313.5	13.4	676 5 Q45553	Q45553 caenorhabdi
13	313.5	13.4	782 5 Q8I412	Q8I412 caenorhabdi
14	292.5	12.5	643 11 Q8VHA7	Q8VHA7 mus musculus
15	283	12.1	363 4 Q8NEC6	Q8NEC6 homo sapien
16	272.5	11.6	278 11 Q8VBX0	Q8VBX0 mus musculus

17	266.5	11.4	329	11	Q9DIA4	Q9DIA4 mus musculus
18	261.5	11.2	276	4	Q8N7B5	Q8N7B5 homo sapien
19	258	11.0	603	11	Q9QV59	Q9QV59 mus musculus
20	258	11.0	603	11	Q9D4L1	Q9D4L1 mus musculus
21	253.5	10.8	355	11	Q8CD00	Q8CD00 mus musculus
22	249.5	10.6	284	11	Q8CCB6	Q8CCB6 mus musculus
23	238.5	10.2	2027	13	Q8UW52	Q8UW52 fugu rubrip
24	237	10.1	1599	11	Q99NH0	Q99NH0 mus musculus
25	235	10.0	435	4	Q9HA95	Q9HA95 homo sapien
26	235	10.0	616	4	Q9EGK0	Q9EGK0 homo sapien
27	235	10.0	627	4	Q8WY90	Q8WY90 homo sapien
28	235	10.0	1486	4	Q8TEF1	Q8TEF1 homo sapien
29	235	10.0	2542	4	Q8IWZ3	Q8IWZ3 homo sapien
30	235	10.0	2617	4	Q8IWZ2	Q8IWZ2 homo sapien
31	233.5	10.0	525	11	Q9CWK6	Q9CWK6 mus musculus
32	233	9.9	627	4	Q9EG77	Q9EG77 homo sapien
33	231	9.9	833	4	Q9E186	Q9E186 homo sapien
34	231	9.9	1188	4	Q9H288	Q9H288 homo sapien
35	230.5	9.8	525	11	Q8R5B4	Q8R5B4 mus musculus
36	228	9.7	4001	5	Q8WR07	Q8WR07 drosophila
37	224.5	9.6	321	5	Q9VQI1	Q9VQI1 drosophila
38	224.5	9.6	619	5	Q9VM19	Q9VM19 drosophila
39	224.5	9.6	1719	4	Q13768	Q13768 homo sapien
40	224.5	9.6	1848	11	Q61302	Q61302 mus musculus
41	224.5	9.6	1856	4	Q99407	Q99407 homo sapien
42	224	9.6	323	11	Q9CQ31	Q9CQ31 mus musculus
43	223.5	9.5	1136	6	Q9N180	Q9N180 bos taurus
44	223	9.5	1181	5	Q9XZ37	Q9XZ37 drosophila
45	223	9.5	1181	5	Q9VBP3	Q9VBP3 drosophila

ALIGNMENTS

RESULT 1

ID Q8N8F5 PRELIMINARY; PRT; 318 AA.
AC Q8N8F5;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ39577.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RA Tachiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
RA Isogai T.;
RT "NEO human cDNA sequencing project";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK096896; BAC04888.1; -
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 5.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 5.
DR PROSITE; PS50088; ANK_REPEAT; 3.
DR PROSITE; PS50297; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REPEAT; Repeat.
KW Hypothetical protein; ANK repeat; Repeat.
SQ SEQUENCE 318 AA; 34980 MW; C4D8F57AF22E629F CRC64;

Query Match 69.8%; Score 1636; DB 4; Length 318;
Best Local Similarity 99.7%; Pred. No. 1.1e-137;
Matches 317; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 132 MRIVLQAKWNLMDITKIFSLQPDKEEDTDTTEKQALNQAVYDNDSTYLLDQLLRQERY 191
DB 1 MRIVLQAKWNLMDITKIFSLQPDKEEDTDTTEKQALNQAVYDNDSTYLLDQLLRQERY 60
QY 192 KRFINSRSGWGPVGTPLRLAASVGHLSCLQVLLAHGADVDSLDVKAQTPLFTAVSHGHLD 251
DB 61 KRFINSRSGWGPVGTPLRLAASVGHLSCLQVLLAHGADVDSLDVKAQTPLFTAVSHGHLD 120
QY 252 CVRVLLAEAGSPGSGSYNNCSPLVTAARDGAVAILQELLHGAENAVKALPVMASNIAS 311
DB 121 CVRVLLAEAGSPGSGSYNNCSPLVTAARDGAVAILQELLHGAENAVKALPVMASNIAS 180
QY 312 CSGPLYAAVGHLDLQVLLAHGADVDSLDVKAQTPLFTAVSHGHLD 371
DB 181 CSGPLYAAVGHLDLQVLLAHGADVDSLDVKAQTPLFTAVSHGHLD 240
QY 372 QLLIDFGANILFSLDLTSDQDKGIALLOARATPRSLLSQVRLVVRALCOAQPOA 431
DB 241 QLLIDFGANILFSLDLTSDQDKGIALLOARATPRSLLSQVRLVVRALCOAQPOA 300
QY 432 INQLDIPPMILSYLKHQ 449
DB 301 INQLDIPPMILSYLKHQ 318

RESULT 2

Q9D738 PRELIMINARY; PRT; 308 AA.
AC Q9D738;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE 2310036C05Rik protein (Ankyrin repeat domain-containing SOCS box
protein Asb-12).
GN 2310036C05RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Giasi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayaishizaki Y.
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690 (2001).
RN [2]
SEQUENCE FROM N.A.
RP Kile B.T., Nicola N.A.;
RT "SOCS box proteins";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK009644; BAB26410.1; -;
DR EMBL; AF403040; AAL57359.1; -;
DR MGD; MGI:1917642; 2310036C05Rik.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 5.

DR PRINTS; PRO1415; ANKYRIN.
DR SMART; SM00248; ANK; 5.
DR PROSITE; PS00088; ANK_REPEAT; 3.
DR PROSITE; PS0297; ANK_REPEAT; 1.
KW ANK repeat; Repeat.
SQ SEQUENCE 308 AA; 33911 MW; 498246E6616DB961 CRC64;

Query Match 61.2%; Score 1435; DB 11; Length 308;
Best Local Similarity 88.6%; Pred. No. 9.7e-120; Indels 0; Gaps 0;
Matches 273; Conservative 17; Mismatches 18;

QY 141 MNLMIDITKIFSLQPDKEEDTDTTEKQALNQAVYDNDSTYLLDQLLRQERYKRFINSRG 200
DB 1 MNLMIDITKIFSLQPDKEEDTDTTEKQALNQAVYDNDSTYLLDQLLRQERYKRFINSRG 60
QY 201 MCVPTPLRLAASVGHLSCLQVLLAHGADVDSLDVKAQTPLFTAVSHGHLD 260
DB 61 MCVPTPLRLAASVGHLSCLQVLLAHGADVDSLDVKAQTPLFTAVSHGHLD 120
QY 261 ASPGSIYNNCSPLVTAARDGAVAILQELLHGAENAVKALPVMASNIAS 320
DB 121 ASPGSIYNNCSPLVTAARDGAVAILQELLHGAENAVKALPVMASNIAS 180
QY 321 VYGHLDLQVLLAHGADVDSLDVKAQTPLFTAVSHGHLD 380
DB 181 VYGHLDLQVLLAHGADVDSLDVKAQTPLFTAVSHGHLD 240
QY 381 IYLPSSLDLTSQDDKGIALLQARATPRSLLSQVRLVVRALCOAQPOA 440
DB 241 IYLPSSLDLTSQDDKGIALLQARATPRSLLSQVRLVVRALCOAQPOA 300
QY 441 LISYLKHQ 448
DB 301 LISYLKHQ 308

RESULT 3

Q96EF0 PRELIMINARY; PRT; 704 AA.
ID Q96EF0
AC Q96EF0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to hypothetical protein FLJ20126.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC012399; AAH12399.1; -;
DR Genew; HGNC:16825; MTMR8.
DR InterPro; IPR000387; TYR_phosphatase.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
KW Hydrolase.
SQ SEQUENCE 704 AA; 78918 MW; 515BE817C9AEA961 CRC64;

Query Match 28.8%; Score 675.5; DB 4; Length 704;
Best Local Similarity 38.7%; Pred. No. 2.3e-51;
Matches 186; Conservative 44; Mismatches 128; Indels 123; Gaps 18;

QY 1 MRESGKLIIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPPEIVVPKSVTLGTVGSSK 60
DB 122 MRESGKLIIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPPEIVVPKSVTLGTVGSSK 181
QY 61 FRSKERVPVLSYLYKENNAIICRCQPLSGFTTRCVDDLELLEASQTNPQSGQFMVVD 120
DB 182 FRSKERVPVLSYLYKENNAIICRCQPLSGFTTRCVDDLELLEASQTNPQSGQFMVVD 241
QY 121 RPKI-----WHPL-----VLIMRIVLQAKWNLMDITKIFSLQ 154

```
|||||
242 RPKLNAMANRAAGKGYENEDNYANIRFRPMGTENIHVWRSLQ-----KLEVECELKT 294
QY 155 PKKEEDTTEBKQALN--QAVYDNDSTYLDOLLQOEYKRFINSRSGWVGPTPLRLAA 212
Db 295 PTMSEFLSGLESSGWLRIKAIMDAGIF-ITKAVKVKASVLVHCSGDGMDRTAQCVCSVAS 353
QY 213 -----SYGHLSCLOVLL-----AHGADVSLDVKAQTPLFTAVSHGHLD 252
Db 354 ILDDPPYRTFKGLMILIEKEMISMGHKFSQRCGH-LDGDSEK-----SPIFTQ-----FLDC 405
QY 253 VRVLE-----AGASPGGSIYNNCSFVLTAARDGNAVILQELLHNG 293
Db 406 IQLMEOQFFCAPEFNFENFLLEIHDHFVSCQFNGFLNCO-----KOREDLR 451
QY 294 AEANVKAKLPVWASNTASCSGPLYLA-AVYGHLD---CFRLLLLHGADPDVNCYDQGL- 347
Db 452 VYEKTHSVWFFLVQRKDPFRNPLVYKGFYVGLNFPSTVPYIQFCWGM---YNRFDKGLQ 508
QY 348 -----LARVPRPTLLLEICLH-----HNCPEYIQLLIDFGANIYLPSSLDT 391
Db 509 PKQSMLESLLLEIKQRAMLETDVHELEKLVYRDEPPEIEICTCSQLG-NILSGLGSLPT 567
QY 392 S 392
Db 568 N 568

RESULT 4
QY3N3 PRELIMINARY; PRT; 401 AA.
AC QY3N3;
DT 01-NOV-1999 (TREMELrel. 12, Created)
DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE DJ710L4.2 (Similar to MYOTUBULARIN-related protein) (Fragment).
GN DJ710L4.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chapman J.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL034408; CAB38778.1; -.
DR InterPro; IPR000387; TYR_phosphatase.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
KW Hydrolase.
FT NON_TER 1
FT NON_TER 401
FT NON_TER 401
SQ SEQUENCE 401 AA; 45607 MW; 8044A4031B35298C CRC64;

Query Match 28.7%; Score 673; DB 4; Length 401;
Best Local Similarity 50.7%; Pred. No. 1.7e-51;
Matches 155; Conservative 28; Mismatches 59; Indels 64; Gaps 10;

QY 1 MRESGKLLDPIISDFGRMGIPNRNWTITDANRNYEICTSTYPPEIVVPKSVTLGTWVGSSK 60
Db 73 MRESGKLLDPIISDFGRMGIPNRNWTITDANRNYEICTSTYPPEIVVPKSVTLGTWVGSSK 132
QY 61 PFSKERVPLSYLYKENNAACRCQPLSGFYTRCVDDLELLEASQTNPGSQFMYVVD 120
Db 133 PFSKERVPLSYLYKENNAACRCQPLSGFYTRCVDDLELLEASQTNPGSQFMYVVD 192
QY 121 RPKI-----WHFL-----VLIMRIVLQAKMNLMDITKIFSLIQ 154
Db 193 RPKLNAMANRAAGKGYENEDNYANIRFRPMGTENIHVWRSLQ-----KLEVECELKT 245
QY 155 PKKEEDTTEBKQALN--QAVYDNDSTYLDOLLQOEYKRFINSRSGWVGPTPLRLAA 212
Db 246 PTMSEFLSGLESSGWLRIKAIMDAGIF-ITKAVKVKASVLVHCSGDGMDRTAQCVCSVAS 304
```

```
213 -----SYGHLSCLOVLL-----AHGADVSLDVKAQTPLFTAVSHGHLD 252
Db 305 ILDDPPYRTFKGLMILIEKEMISMGHKFSQRCGH-LDGDSEK-----SPIFTQ-----FLDC 356
QY 253 VRVLE 258
Db 357 IWQLME 362

RESULT 5
QY3N3 PRELIMINARY; PRT; 327 AA.
AC QY3N3;
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE 1700023008Rik protein.
GN ASB1 OR 1700029008RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Mateu Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tonita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Butt C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 403:685-690(2000).
DR EMBL; AK006515; BAB24628.1; -.
DR HSSP; Q00420; IAWC.
DR MGD; MGI:1929735; Asb1.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 6.
DR SMART; SM00248; ANK; 5.
DR PROSITE; PS00088; ANK_REPEAT; 3.
DR PROSITE; PS0297; ANK_REPEAT; 1.
KW ANK repeat; Repeat.
SQ SEQUENCE 327 AA; 36281 MW; CB0649F3D080910C CRC64;

Query Match 20.5%; Score 480.5; DB 11; Length 327;
Best Local Similarity 40.7%; Pred. No. 2e-34;
Matches 111; Conservative 51; Mismatches 90; Indels 21; Gaps 7;

QY 170 LNAQVYDNDSTYLDOLLQOEYKRFINSRSGW---GVGFTPLRLAASVYCHLSCLQVLLAH 286
Db 34 LHDAAYVGDQLTRLNLLQEEYSRISRNKSVCCWGLPCTPLRIATAATAGHCNVDLIRK 93
QY 227 GADVSLDVKAQTPLFTAVSHGHLDVRLLEAGASPGGSIYNNCSFVLTAARDGNAVIL 286
Db 94 GAEDVLVDVKGQALYVAVVNGHLESTELLEAGADPNGRHRSRTPYVHASRVGRDDL 153
QY 287 QELDLHGAEANVYAKL-----PVWASNTAS---CSGPLYLAAYVGHLCDFRLLHGD 338
Db 154 KALIRYAGADVNVNHLTPDTRPPFSRRRLTSLVLC--PLYISAAYHNLQCFRLLQAGANP 211
```

```

QY 339 DYNCTDQGLLARVR--PRTLLEICLHNCCEPEYIOLLIDFGANIYL---PSISLDTLSQ 393
Db 212 DFNCNGPNTQEFYRSPGCVDAVRHGCCEAAVSLVFEFGANLNLVKWESLGPARGR 271
QY 394 ---DDKGIALQLQARATPRSLISQVRLVRRAL 423
Db 272 RKWDPEALQVFEKARSIPRTLISLCRVAVRRAL 304

RESULT 6
Q8VE11 PRELIMINARY; PRT; 617 AA.
AC Q8VE11;
DT 01-WAR-2002 (T-EMBLrel. 20, Created)
DT 01-WAR-2003 (T-EMBLrel. 20, Last sequence update)
DT 01-WAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical protein (MYOTUBULARIN related protein 6 homolog).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; BC020019; AAH20019.1; -.
DR EMBL; AK076218; BAC36259.1; -.
DR InterPro; IPR000387; TYR_phosphatase.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1;
KW Hypothetical protein; Hydrolase.
SQ SEQUENCE 617 AA; 70932 MW; EE699C578602A013 CRC64;

Query Match 18.0%; Score 421; DB 11; Length 617;
Best Local Similarity 59.3%; Pred. No. 1e-28;
Matches 73; Conservative 25; Mismatches 25; Indels 0; Gaps 0;

QY 2 RESGWKLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPPIVVPKSVTLGTWVGSSKF 61
Db 121 RRGWQLDLAAEYERMGVNPANWQLSDANREYKVCETYPRELYVPTASRPVIVGSSNF 180
QY 62 RSKERVPLSYLYKENNAICRCSQPLSGFYTRCVDDLELLAISQTNPGSQFMVVDTR 121
Db 181 RSKGRPLVLSYCRQGTAAICRCSQPLSGFSARCLEDEHLLQAISKANPNRYMYVVDTR 240

Query Match 16.9%; Score 396; DB 11; Length 467;
Best Local Similarity 55.3%; Pred. No. 1.2e-26;
Matches 68; Conservative 25; Mismatches 30; Indels 0; Gaps 0;

QY 2 RESGWKLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPPIVVPKSVTLGTWVGSSKF 61
Db 123 REQWLLVLDLSEYKRMGLPDNYWQLSDVNRDRCVDSYPTLYVPRSATAHIIVGSSKF 182
QY 62 RSKERVPLSYLYKENNAICRCSQPLSGFYTRCVDDLELLAISQTNPGSQFMVVDTR 121
Db 183 RSRRPPLSYKCKSHASICKSSQPLSGFSARCLEDEHLLQAISKANPNRYMYVVDTR 242
QY 122 PKI 124
Db 243 PKL 245

RESULT 7
Q96P80 PRELIMINARY; PRT; 621 AA.
AC Q96P80;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
DE Myotubularin related protein 6.
GN MTMR6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Hong W.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF408619; AAL01037.1; -.
DR InterPro; IPR000387; TYR_phosphatase.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
KW Hydrolase.
SQ SEQUENCE 621 AA; 71896 MW; A7346D3AE82911A7 CRC64;

Query Match 17.4%; Score 408; DB 4; Length 621;
Best Local Similarity 58.5%; Pred. No. 1.5e-27;
Matches 72; Conservative 25; Mismatches 26; Indels 0; Gaps 0;

QY 2 RESGWKLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPPIVVPKSVTLGTWVGSSKF 61
Db 121 RLQHQQLIDLAEYKRMGVNPNSHWQLSDANRDYKICETYPRELYVPRIASKPIIVGSSKF 180
QY 62 RSKERVPLSYLYKENNAICRCSQPLSGFYTRCVDDLELLAISQTNPGSQFMVVDTR 121
Db 181 RSKGRFPVLSYHYHDKKAAICRCSQPLSGFSARCLEDEHLLQAISKANPNRYMYVVDTR 240
QY 122 PKI 124
Db 241 PKL 243

RESULT 8
Q8K0A2 PRELIMINARY; PRT; 467 AA.
AC Q8K0A2;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-WAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Similar to myotubularin related protein 6.
GN MTMR7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC032254; AAH32254.1; -.
DR MGI; MGI:1891693; Mtmr7.
DR InterPro; IPR0003595; PTPC_motif.
DR InterPro; IPR000387; TYR_phosphatase.
DR SMART; SM00404; PTPC_motif; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
SQ SEQUENCE 467 AA; 53927 MW; F756FB90E316A9A5 CRC64;

Query Match 16.9%; Score 396; DB 11; Length 467;
Best Local Similarity 55.3%; Pred. No. 1.2e-26;
Matches 68; Conservative 25; Mismatches 30; Indels 0; Gaps 0;

QY 2 RESGWKLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPPIVVPKSVTLGTWVGSSKF 61
Db 123 REQWLLVLDLSEYKRMGLPDNYWQLSDVNRDRCVDSYPTLYVPRSATAHIIVGSSKF 182
QY 62 RSKERVPLSYLYKENNAICRCSQPLSGFYTRCVDDLELLAISQTNPGSQFMVVDTR 121
Db 183 RSRRPPLSYKCKSHASICKSSQPLSGFSARCLEDEHLLQAISKANPNRYMYVVDTR 242
QY 122 PKI 124
Db 243 PKL 245

RESULT 9
Q8C4J6 PRELIMINARY; PRT; 660 AA.
ID Q8C4J6
AC Q8C4J6;

```

```

DT 01-MAR-2003 (trEMBLrel. 23, Created)
DT 01-MAR-2003 (trEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (trEMBLrel. 23, Last annotation update)
DE Myotubularin related protein 7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA THE FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK081973; BAC38383.1; -.
SQ SEQUENCE 660 AA; 75608 MW; 57CA1A447192A52B CRC64;

Query Match 16.9%; Score 396; DB 11; Length 660;
Best Local Similarity 55.3%; Pred. No. 2e-26;
Matches 68; Conservative 25; Mismatches 30; Indels 0; Gaps 0;

QY 2 RESGKWLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPETVVPKSVTLGTWVGSSKF 61
Db 123 REQGLLVLSSEYKEMGLPDNYWQLSDVNRDYRVCDSYPTELYVPRSAHIIVGSSKF 182

QY 62 RSKERVPLSVLYLKENNAICRCSPLSGFYTRCVDDDELLLEAISOTNPGSQPMYVVDTR 121
Db 183 RSRRRPPALSYCKDSHASICRSPQLSGFSARCLEQMLQAIKANKPGSDFIYVVDTR 242

QY 122 PKI 124
Db 243 PKL 245

RESULT 10
Q9W1Q6 PRELIMINARY; PRT; 676 AA.
AC Q9W1Q6; Q9W3A1;
DT 01-WAY-2000 (trEMBLrel. 13, Created)
DT 01-WAY-2000 (trEMBLrel. 13, Last sequence update)
DT 01-WAY-2000 (trEMBLrel. 23, Last annotation update)
DE BCDNA:GH04637 protein.
GN BCDNA:GH04637 OR C3530 OR CG18093.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

```

```

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Stradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RC STRAIN=BERKELEY; TISSUE=Head;
RX MEDLINE=20196012; PubMed=10731138;
RA Rubin G.M., Hong L., Brokstein P., Evans-Holm M., Frise E.,
RA Stapleton M., Harvey D.A.;
RT "A Drosophila complementary DNA resource."
Science 287:2222-2224 (2000).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
DR EMBL; AE003460; AAF46997.1; -.
DR EMBL; AE003460; AAF46999.1; -.
DR EMBL; AF181629; AAD55415.1; -.
DR Flybase; FBgn0028497; BCDNA:GH04637.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR00387; TYR phosphatase.
DR InterPro; IPR00306; Znf_FYVE.
DR Pfam; PF01363; FYVE; 1.
DR SMART; SM00064; FYVE; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00383; TYR PHOSPHATASE_1; 1.
DR PROSITE; PS00566; TYR PHOSPHATASE_2; 1.
DR PROSITE; PS0178; ZF_FYVE; 1.
KW Hypothetical protein, Alternative splicing; Hydrolase.
FT VARSPPLIC 1 37 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 676 AA; 77223 MW; 988BE4C052754B43 CRC64;

Query Match 15.4%; Score 360; DB 5; Length 676;
Best Local Similarity 26.9%; Pred. No. 3.4e-23;
Matches 119; Conservative 62; Mismatches 159; Indels 102; Gaps 15;

QY 2 RESGKWLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPETVVPKSVTLGTWVGSSKF 61
Db 108 KNAGWDYFKLEAEFKHMLVNEAWTLCSMNEKYELCDYTPRQIYVPKEATLTLLISSRF 167

QY 62 RSKERVPLSVLYLKENNAICRCSPLSGFYTRCVDDDELLLEAISOTNPGSQPMYVVDTR 121
Db 168 RSKGRPLVLTLYLH-NNKASICRCSPLSGFSARCLEQMLLEAIRKNTNSNTDYMVVDTR 226

QY 122 PKI-----WHFVLVIMRVLQAKNMLMDITKIFSLQPDKEE 159
Db 227 PRINAMNPAACKGYNEAFYENIKHFILG-IENIHVQASLQ-----KVLEACEQK 277

QY 160 EDTDTEEKQALN-----QAVYNDSTYTLQDLQERYKRFINSRSGVPG-----T 206
Db 278 SPTMSAFINALESGWLKHRSILDTSSFIANAV--DKGVSVVHCHSDGMDRTAQVCSLA 335

QY 207 PRLAASYGHLSCQLV-----LAHCADVD-----SLDVKAQTPLTFTAVSHGLDCV-- 253
Db 336 QLMNLPYRTIKGFQALIEKDWLAFGHKFSRCGHITQTDAREVSPFTQ-----FLDCTWQ 391

QY 254 -----RVLL-----EAGSPGSGIYNNCSPLVTAARDGAVAILQELLDHGAEA 296

```

```

Db 392 LMSORSEAFENRFLILLHDHVHSCQFGTFVGNCE-----KORLDLKLAE 437
Qy 297 NVKAKLPVWASNTASCGPLYL-----AAVYGHLDCEFRLLLLHGADPDYNTDQGLLARVP 352
Db 438 RTFSLMGYMANHLEYNIPKPNVDIAKANLAPQCIKWGRM---YSFESGIHPREP 494
Qy 353 RPTLLIEICLHNCPEPIYQILL 374
Db 495 LGVDLLDSKHCNSLEDHVQHL 516

RESULT 11
Q8MLR7
ID Q8MLR7 PRELIMINARY; PRT; 761 AA.
AC Q8MLR7;
DT 01-OCT-2002 (T8EMBLrel. 22, Created)
DT 01-OCT-2002 (T8EMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (T8EMBLrel. 23, Last annotation update)
DE C83530-PA.
GN BCDNA:GH04637 OR CG3530 OR CG18093.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophiliidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely; PubMed=107311132;
RX MEDLINE=20196006;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Franknoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Riden-Kiamos I., Simpson M., Skupski M.P., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Rector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong P.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Ananides P.G., Brandon R.C., Rogers Y.,
RA Banton J., An H., Baldwin D., Banton J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,

```

```

RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dreesnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradscky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clump W., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003460; AAF46998.2; -.
DR FlyBase; FBgn028497; BCDNA:GH04637.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003595; PTPC motif.
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000306; Znf_FYVE.
DR Pfam; PF01363; FYVE; 1.
DR SMART; SM00064; FYVE; 1.
DR SMART; SM00404; PTPC motif; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00556; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS0178; ZF_FYVE; 1.
SQ SEQUENCE 761 AA; 87047 MW; 11EED6502933A860 CRC64;

```

```

Query Match 15.4%; Score 360; DB 5; Length 761;
Best Local Similarity 26.9%; Pred. No. 4e-23;
Matches 119; Conservative 62; Mismatches 159; Indels 102; Gaps 15;
Qy 2 RESGWKLIDPIGDFGRMGIPNRRNWTITDANRNYEICSTYPPPIVVPKSVTLGTWVGSSKF 61
Db 193 KNAGWDYFKLEAFKHLVPNEAWTICSMNKEYELCDYTPROIYVPKENTLMLISSRF 252
Qy 62 RSKRPVPLSVLYKENNAICRCSQPLSGFYTRCVDDLELLLEAISQTFGSMYVVDTR 121
Db 253 RSKGRPLVLYLH-NKASICRCSQPLSGFSARCLEDQMLEAIRKTSNTDYMVVDTR 311
Qy 122 PKI-----WHFLVLMRIVLQAKVLMMDITKIFSLQPDKEE 159
Db 312 PRINAMANAAGKGYENAFYENIKFHLG-IENIHVQRASLQ-----KVLEACEQK 362
Qy 160 EDTDTEERQALN-----QAVYNDSDVTLDQLLRQERYKRFINSRSGVPG-----T 206
Db 363 SPTMSAFINALLESSGWLKHIRSLDTSFFIANAV--DKGVSVVVHSCSDGWDRTAQVCSLA 420
Qy 207 PLRLAASYGHLSCLQVL-----LAHGADVD-----SLDVKAQTPLFTAVSHGLDCV-- 253
Db 421 QLMNLNPPYRTIKGFQALLIEKDWLAFGHKFSERCIGHIQTDAREVSPIFTQ-----FLDCTWQ 476
Qy 254 -----RVLL-----EAGASPGSGSYNNCSPLVTAARDGAVAILQELLHGABA 296
Db 477 LMSORSEAFENRFLILLHDHVHSCQFGTFVGNCE-----KORLDLKLAE 522

```


Db 263 SRGRIPVLSWHPESQATVTRCSQPMVGSGSKRKEDEKYLQAIMDSNAQSHKIFIFDAR 322
Qy 122 PKI-----WHFLVLMIRIVLQAKNLM--DITKIFSLIQPKKE---159
Db 323 PSVNAVANKAKGGYSESDAYQNAELVF---LDIHNTHVMRESLRKLEIYVPTIEETHW 379
Qy 160 ----EDT-----DTEEKQALNOAVYDNDSY-----TLDQLLRQERYK 192
Db 380 LSNLESTHLEHIKILLAGALRIADKVESGKTSVVVHCSGWDRTAQLTSLAMLMDGY 439
Qy 193 RFINS-----RSGWGVPTPLRLAASYGHLSCLOVLLAHG-----ADVDSL DVKAOTPLFT 243
Db 440 RTIRGEVULVEKEW-----LSFGHR--FQLRVGHGDKKHADAD-----RXEVFL 481
Qy 244 AVSHGHLDVCVRVLLLEAGSPGSIYNNCSPLVTAARDCAVAIIQELLDH-----292
Db 482 Q-----FIDCVWQWTR--QFTAFEFNE-----YFLITILDHLYSLFGTFLC 522
Qy 293 -----GAEANVKAKLPWA-----SNIASCSGPLY 317
Db 523 NSEQQKGKGNLPKKTSLWSYINSQLEDFTNPLY 556

RESULT 15

Q8NEC6 PRELIMINARY; PRT; 363 AA.
AC Q8NEC6;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to myotubularin related protein 1.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC032947; AAH32947.1; -
DR InterPro; IPR004182; GRAM dom.
DR Pfam; PF02893; GRAM; 1.
DR SMART; SM00568; GRAM; 1.
SQ SEQUENCE 363 AA; 39822 MW; 9AE00536BCF60811 CRC64;

Query Match 12.1%; Score 283; DB 4; Length 363;
Best Local Similarity 43.5%; Pred. No. 1e-16;
Matches 57; Conservative 21; Mismatches 45; Indels 8; Gaps 2;
Qy 4 SGWKLIDPISDFGRMGIPNRRNTITDANRNYEICSTVPPPIVVPKSVTLGTVVGSSKPRS 63
Db 225 NGWKVYDPVSEYKRGQLPNESWKISKINSYECDTYPALIVVPTS VKDDDSLKVAAPRA 284
Qy 64 KERVPLSVLYKNNAAICRCSQPLSG-FYTRCVDDLELLEAISOTNPGSQFMVYVDTRP 122
Db 285 KGRVPVLSWHPESQATITRCSQPLVGPNDKCKEKEKYLQTIMDANAQSHKLIIFDARQ 344
Qy 123 -----KIWH 126
Db 345 NSVADTNKLMY 355

Search completed: December 1, 2003, 13:45:30
Job time : 76 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 1, 2003, 09:17:14 ; Search time 6732 Seconds
(without alignments)
11412.400 Million cell updates/sec

Title: US-09-941-831A-6

Perfect score: 1878

Sequence: 1 catgattacgccagcttgg.....taaaaaaaaaaaaaaaaa 1878

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pin.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1119.4	59.6	1269	9	AK096896	AK096896 Homo sapi
2	1117.8	59.5	1222	6	AX418418	AX418418 Sequence
3	931	49.6	1069	6	AX418416	AX418416 Sequence
4	924	49.2	927	9	AF403030	AF403030 Homo sapi
c	851.4	45.3	104534	9	AL356317	AL356317 Human DNA
5	729.2	38.8	1262	10	AF403040	AF403040 Mus muscu
6	710.2	37.8	2653	9	BC012399	BC012399 Homo sapi
7	703.4	37.5	1345	6	AX406008	AX406008 Sequence
8	628.6	33.5	191093	10	AL671765	AL671765 Mouse DNA
c	612	32.6	194079	2	AC109692	AC109692 Rattus no
10	612	32.6	225982	2	AC108644	AC108644 Rattus no
11	290.6	15.5	240816	5	AL844570	AL844570 Zebrafish
12	254.2	13.5	2354	5	BC053411	BC053411 Danio rer
13	241.8	12.9	2353	6	AX086038	AX086038 Sequence
14	238.2	12.7	2754	5	BC044359	BC044359 Danio rer
15	233.6	12.4	2817	10	BC032254	BC032254 Mus muscu
16	228	12.1	3065	10	BC020019	BC020019 Mus muscu
17	224	11.9	1866	9	AF406619	AF406619 Homo sapi
18	224	11.9	2337	9	AK093237	AK093237 Homo sapi
19	224	11.9	4069	9	BC040012	BC040012 Homo sapi
20	224	11.9	4131	9	HSN803324	HSN803324 Homo sapi
21	220.8	11.8	4131	9	AK000133	AK000133 Homo sapi
22	204.4	10.9	1997	9	BT003304	BT003304 Drosophil
23	181.4	9.7	3059	3	HS7101A	HS7101A Human DNA
c	169	9.0	129849	9	AF181629	AF181629 Drosophil
25	149.6	8.0	3209	3	AX398878	AX398878 Sequence
26	135.8	7.2	2482	6	AF155352	AF155352 Mus muscu
27	109.6	5.8	1008	10	AK122456	AK122456 Mus muscu
28	109.6	5.8	5892	10	AK122456	AK122456 Mus muscu
29	109.6	5.8	5892	10	AK122456	AK122456 Mus muscu
30	108.8	5.8	2019	6	AX022304	AX022304 Sequence
31	108.8	5.8	2019	6	AX030771	AX030771 Sequence
32	108.8	5.8	2019	6	AX030771	AX030771 Sequence
33	108.8	5.8	2019	6	BD007458	BD007458 Remedies
34	103.6	5.5	3382	9	AF072928	AF072928 Homo sapi
35	100.8	5.4	1008	9	AF156777	AF156777 Homo sapi
36	100.8	5.4	1340	9	BC014528	BC014528 Homo sapi
37	100.2	5.3	1648	9	AF073482	AF073482 Homo sapi
38	93.6	5.0	6518	9	AB032972	AB032972 Homo sapi
39	92	4.9	190461	2	AC109199	AC109199 Mus muscu
40	92	4.9	228645	2	AC094273	AC094273 Rattus no
41	83	4.4	42724	3	CEP53A2	281546 Caenorhabdi
42	77.4	4.1	93778	9	AC016999	AC016999 Homo sapi
c	77.4	4.1	172990	2	AC011973	AC011973 Homo sapi
44	76.4	4.1	2397	9	BC052990	BC052990 Homo sapi
45	76.4	4.1	3158	9	BC040432	BC040432 Homo sapi

ALIGNMENTS

RESULT 1
AK096896
LOCUS
DEFINITION Homo sapiens cDNA FLJ39577 fis, clone SKMUS200290, weakly similar
to Homo sapiens ASB-1 protein mRNA.
ACCESSION AK096896
VERSION AK096896.1 GI:21756490
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
Tashiro, H., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S.,
Fukuzumi, Y., Fujimori, Y., Komiyama, M., Sugiyama, T., Irie, R.,
PRI 15-JUL-2002

QY	1690	ATTAGTACTCTAAAAACCAACTGTAATTTTCAGTCTCTCCAGGAACCTTATGATGCCTC	1749		
Db	1062	ATTAGTACTCTAAAAACCAACTGTAATTTTCAGTCTCTCCAGGAACCTTATGATGCCTC	1121		
QY	1750	CGAAAACCACTGGGAGCTACGTAGCTGGAGAGCATTTACAGGCTCATCAGCTTACTG	1809		
Db	1122	CGAAAACCACTGGGAGCTACGTAGCTGGAGAGCATTTACAGGCTCATCAGCTTACTG	1181		
QY	1810	AGCTGCTCTCCGTATATTATCTCCCAATAAAAAATTCCTCAG	1850		
Db	1182	AGCTGCTCTCCGTATATTATCTCCCAATAAAAAATTCCTCAG	1222		
RESULT 3					
LOCUS	AX418416	1069 bp	DNA linear PAT 18-JUN-2002		
DEFINITION	Sequence 11 from Patent WO0206329.				
ACCESSION	AX418416				
VERSION	AX418416.1	GI:21523341			
KEYWORDS	unidentified				
SOURCE	unidentified				
ORGANISM	unclassified.				
REFERENCE	1				
AUTHORS	Rastelli, L., Shimkets, R.A., Zehrhusen, B., Malyankar, U.M. and Padigaru, M.				
TITLE	Human polynucleotides and polypeptides encoded thereby				
JOURNAL	Patent: WO 0206329-A 11 24-JAN-2002;				
FEATURES	Curagen Corporation (US)				
source	Location/Qualifiers				
	1..1069				
	/organism="unidentified"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:32644"				
	/note="NOVX nucleic acid with homology to ankyrin repeat-containing- (ASB-1) like protein"				
BASE COUNT	237 a	298 c	257 g		
ORIGIN					
Query Match	49.6%; Score 931; DB 6; Length 1069;				
Best Local Similarity	97.6%; Pred. No. 1.2e-254;				
Matches	965; Conservative	0; Mismatches	0; Indels		
	24; Gaps	1;			
QY	760	ATGAGATAAGTTCTCCAATTAGCCAAGATGAACCTCATGGACATCACCAAGATCTTCTCC	819		
Db	81	ATGAGATAAGTTCTCCAATTAGCCAAGATGAACCTCATGGACATCACCAAGATCTTCTCC	140		
QY	820	CTCTGCAGCCGACAAAGAGGAGGAGGACATGTACACAGAGAGGAAGCAGGCTCTCAAT	879		
Db	141	CTCTGCAGCCGACAAAGAGGAGGAGGACATGTACACAGAGAGGAAGCAGGCTCTCAAT	200		
QY	880	CAAGCAGTGTATGACAAACACTCTTATCTTTGGACAGCTTTTGGCCAGGAGCGTTAC	939		
Db	201	CAAGCAGTGTATGACAAACACTCTTATCTTTGGACAGCTTTTGGCCAGGAGCGTTAC	260		
QY	940	AAACGTTTCATCAACAGCAGGAGTGGCTGGGGTGTTCTCTGGGACACCTTGCCTTTGGCT	999		
Db	261	AAACGTTTCATCAACAGCAGGAGTGGCTGGGGTGTTCTCTGGGACACCTTGCCTTTGGCT	320		
QY	1000	GCTTCTTATGGCCACTTGAGCTGTTTTCGAAGTCTCTTAGCCCCATGGTCTGATGTGAC	1059		
Db	321	GCTTCTTATGGCCACTTGAGCTGTTTTCGAAGTCTCTTAGCCCCATGGTCTGATGTGAC	380		
QY	1060	AGCTTGGATGTCAAGGCACAGACGCCACTTTTCACTGCTGTCAAGTCAATGGCCATCTGGAC	1119		
Db	381	AGCTTGGATGTCAAGGCACAGACGCCACTTTTCACTGCTGTCAAGTCAATGGCCATCTGGAC	440		
QY	1120	TGTTGACGTGTGCTTTTGAAGCTGTGTGCTCTCTCTGTTGGTAGCATCTTACAACTCTGT	1179		
Db	441	TGTTGACGTGTGCTTTTGAAGCTGTGTGCTCTCTCTGTTGGTAGCATCTTACAACTCTGT	500		
QY	1180	TCTCCGCTGCTCACAGCTGCCCGTGATGGTGTGTTGCTATCTCTGAGAGAGCTCTTAGAC	1239		

Db 501 TCTCCGCTGCTCACAGCTCCCGTGTATGCTGTTGCTATCTCGCAGGAGCTCCTAGAC 560
Qy 1240 CATGTGTCAGAGGCCAAGCTCAAGCTAACTACAGCTGCGGCATCAAACTAGCTTCA 1299
Db 561 CATGTGTCAGAGGCCAAGCTCAAGCTAACTACAGCTGCGGCATCAAACTAGCTTCA 620
Qy 1300 TGTCTCGGCCCTCTATTGTCGCGCAGTCTAGGGCAGCTGAGCTGTTTCGCGCTGCTT 1359
Db 621 TGTCTCGGCCCTCTATTGTCGCGCAGTCTAGGGCAGCTGAGCTGTTTCGCGCTGCTT 680
Qy 1360 TTGCTCCAGGGGCGAGCCCTGACTACAACTGCACTGACAGGGCCCTATTGGCTGCTGTC 1419
Db 681 TTGCTCCAGGGGCGAGCCCTGACTACAACTGCACTGACAGGGCCCTATTGGCTGCTGTC 740
Qy 1420 CCAAGACCCGCGACCCCTCTGAAATCTGCTCCATCATATAATGTCAGCAGATATATC 1479
Db 741 CCAAGACCCGCGACCCCTCTGAAATCTGCTCCATCATATAATGTCAGCAGATATATC 800
Qy 1480 CAGCTGTTAATCGATTTGCTGCTAATATCTACCTTCCATCTCTCCCTTGACCTGACC 1539
Db 801 CAGCTGTTAATCGATTTGCTGCTAATATCTACCTTCCATCTCTCCCTTGACCTGACC 860
Qy 1540 TCACAAGATGATAAAGGCAATGCTGCTGCTACAGGCCCG----- 1580
Db 861 TCACAAGATGATAAAGGCAATGCTGCTGCTACAGGCCCGAGGTGAGCTGTTCTCTT 920
Qy 1581 -----AGCACCTCCAGGTCATCTTATCACAGGTCGTTAGTCTCGCAGAGCTTG 1635
Db 921 GCTGTAGCCACTCCAGGTCATCTTATCACAGGTCGTTAGTCTCGCAGAGCTTG 980
Qy 1636 TGCCAGGCTGGCCAGCCACCAAGCCATCAACAGCTGATATTCCTCCCATGTTGATTAGC 1695
Db 981 TGCCAGGCTGGCCAGCCACCAAGCCATCAACAGCTGATATTCCTCCCATGTTGATTAGC 1040
Qy 1696 TACCTAAAAACCAACTGTAATCTTGAC 1724
Db 1041 TACCTAAAAACCAACTGTAATCTTGAC 1069

RESULT 4
AF403030 927 bp mRNA linear PRI 02-JAN-2002
LOCUS Homo sapiens ankryrin repeat domain-containing SOCS box protein
DEFINITION Asb-12 mRNA, complete cds.
ACCESSION AF403030
VERSION AF403030.1 GI:18034079
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 927)
AUTHORS Kile, B.T. and Nicola, N.A.
TITLE SOCS box proteins
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 927)
AUTHORS Kile, B.T. and Nicola, N.A.
TITLE Direct Submission
JOURNAL Submitted (27-JUL-2001) Cancer & Hematology, The Walter & Eliza Hall Institute of Medical Research, Royal Melbourne Hospital, Parkville, Victoria 3050, Australia
FEATURES
source Location/Qualifiers
1. .927
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1. .927
/codon_start=1
/product="ankryrin repeat domain-containing SOCS box protein Asb-12"
/protein_id="AAL57349.1"
/db_xref="GI:18034080"

/translation="WNLMIDITKIFSLQLQDEEDTDEEKQALNQAVYDNDSTYLDQ
LLRBERYKRFINSRSGWPGTPLRLAASVYGHLSCLQVLAAHAGDLSLDVKAQPLF
TAVSHGLDVCVRVLLLEAGASPGGSVLYNNCSPLVTAARDGAVAILDELIDHGAENVKA
KLPVWASNIASCSPLYLAAVYGHLDLCPRLILLHAGADPDYNTDGLLARVPRPTLL
EICLHNCEPBYIQLLIDFGANILPSLSLDTSDQDDKIGALLQARATPRSLLSQVR
LWVRALCOAGPOQAINQLDIPMLISYLKHQ"
BASE COUNT 203 a 274 c 216 g 234 t
ORIGIN
Query Match 49.2%; Score 924; DB 9; Length 927;
Best Local Similarity 100.0%; Pred. No. 1.2e-252;
Matches 924; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 787 ATGAACCTCATGAGCATCACCAAGATCTTCTCCCTCTGCGCCGACGAGGAG 846
Db 1 ATGAACCTCATGAGCATCACCAAGATCTTCTCCCTCTGCGCCGACGAGGAG 60
Qy 847 GACACTGACACAGAGGAGAGCAGGCTCTCAATCAAGCAGTGTATGACAAACGACTCTAT 906
Db 61 GACACTGACACAGAGGAGAGCAGGCTCTCAATCAAGCAGTGTATGACAAACGACTCTAT 120
Qy 907 ACTTTGGACAGCTTTTGGCCGACGAGGCTTACAAACGTTTCAACAGCAGGAGTGGC 966
Db 121 ACTTTGGACAGCTTTTGGCCGACGAGGCTTACAAACGTTTCAACAGCAGGAGTGGC 180
Qy 967 TGGGGTGTCTTGGGACACCCCTTGGCTGCTTCTTATGGCCACTTTGAGCTGTTTG 1026
Db 181 TGGGGTGTCTTGGGACACCCCTTGGCTGCTTCTTATGGCCACTTTGAGCTGTTTG 240
Qy 1027 CAAAGTCTCTTAGCCCATGCTGCTGATGTGACAGCTTTGGATGTCAAGGCACAGACGCCA 1086
Db 241 CAAAGTCTCTTAGCCCATGCTGCTGATGTGACAGCTTTGGATGTCAAGGCACAGACGCCA 300
Qy 1087 CTTTTCACGTGTCAGTCATGCGCCATCTGACACTGTGACGTGCTTTTGAAGCTGCT 1146
Db 301 CTTTTCACGTGTCAGTCATGCGCCATCTGACACTGTGACGTGCTTTTGAAGCTGCT 360
Qy 1147 GCCTCTCTGTTGGTAGCATCTACAACTGTTCTCCGCTGCTCACAGCTGCCGCTGAT 1206
Db 361 GCCTCTCTGTTGGTAGCATCTACAACTGTTCTCCGCTGCTCACAGCTGCCGCTGAT 420
Qy 1207 GGTGCTGTTGCTATCTCTGAGGAGCTCTTAGACCATGTTGTCAGAGGCCAACGCTCAAAGCT 1266
Db 421 GGTGCTGTTGCTATCTCTGAGGAGCTCTTAGACCATGTTGTCAGAGGCCAACGCTCAAAGCT 480
Qy 1267 AAACCTACAGCTGCGGCATCAAACTAGCTTCTGCGCCCTCTATTTGGCCGCA 1326
Db 481 AAACCTACAGCTGCGGCATCAAACTAGCTTCTGCGCCCTCTATTTGGCCGCA 540
Qy 1327 GTCTACGGGCACTGGAGCTGTTTCGCTGCTTTCCTCCAGGGGCGAGCCCTGACTAC 1386
Db 541 GTCTACGGGCACTGGAGCTGTTTCGCTGCTTTCCTCCAGGGGCGAGCCCTGACTAC 600
Qy 1387 AAACCTACAGCTGCGGCATCAAACTAGCTTCTGCGCCCTCTATTTGGCCGCA 1446
Db 601 AAACCTACAGCTGCGGCATCAAACTAGCTTCTGCGCCCTCTATTTGGCCGCA 660
Qy 1447 TGCCTCCATCATATAATGTCAGCAGATATATCCAGCTGTTAATCGATTTTGGTGTAT 1506
Db 661 TGCCTCCATCATATAATGTCAGCAGATATATCCAGCTGTTAATCGATTTTGGTGTAT 720
Qy 1507 ATCTACCTTCCATCTCTCTCCCTGACCTGACCTCACAGATGATAAAGGCAATTCGATTG 1566
Db 721 ATCTACCTTCCATCTCTCTCCCTGACCTGACCTCACAGATGATAAAGGCAATTCGATTG 780
Qy 1567 CTGCTACAGGCCCGGAGCCACTCCAGGTCACCTTCTATCACAGGTCGTTTGTAGTCCGCGC 1626
Db 781 CTGCTACAGGCCCGGAGCCACTCCAGGTCACCTTCTATCACAGTCCGTTTGTAGTCCGCGC 840
Qy 1627 AGAGCTTGTGTCAGGCTGGCCAGCCACAGCCATCAACAGCTGATATTCCTCCCATG 1686
Db 841 AGAGCTTGTGTCAGGCTGGCCAGCCACAGCCATCAACAGCTGATATTCCTCCCATG 900

QY 1687 TTGATTAGCTACCTAAACACCAA 1710
 |||||
 Db 901 TTGATTAGCTACCTAAACACCAA 924

RESULT 5

AL356317/c

LOCUS

DEFINITION

AL356317 104534 bp DNA linear PRI 02-FEB-2001

Human DNA sequence from clone Rp11-284B18 on chromosome Xq12-13.1,

complete sequence.

ACCESSION

AL356317

VERSION

AL356317.8 GI:12666252

KEYWORDS

HTG.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

Lawlor, S.

JOURNAL

Direct Submission

Submitted (02-FEB-2001) Sanger Centre, Hinxton, Cambridgehire,

CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Feb 5, 2001 this sequence version replaced gi:12578192.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence has been finished according to sequence map criteria

as follows. An attempt is made to resolve all sequencing problems,

such as compressions and repeats, but not necessarily within known

annotated repeat sequence elements. Where the sequence is

ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C/elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome X, constructed by the Sanger Centre Chromosome X Mapping

Group. Further information can be found at

http://www.sanger.ac.uk/HGP/ChrX

Rp11-284B18 is from the library RPC1-11.1 constructed by the group

of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone

Rp11-284B18 It may be shorter because we sequence overlapping

sections only once, except for a 100 base overlap.

The true left end of clone Rp11-284B18 is at 1 in this sequence.

The true left end of clone Rp4-710L4 is at 104435 in this sequence.

The true right end of clone Rp11-403E24 is at 63843 in this

sequence.

FEATURES

source

1. .104534

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosomes="X"

/map="q12-13.1"

/clone="Rp11-284B18"

/clone_lib="RPC1-11.1"

1. .4482

/notes="HERVL repeat: matches 1. .4542 of consensus"

4489. .5037

/note="MLT2B repeat: matches 1. .438 of consensus"

5038. .5133

/notes="LMA7 repeat: matches 5960. .6047 of consensus"

5329. .5388

/notes="L1 repeat: matches 2689. .2746 of consensus"

6136. .6274

repeat_region /note="MIR repeat: matches 49. .203 of consensus"
 6859. .7220
 /note="MLT1A1 repeat: matches 1. .365 of consensus"
 8549. .8744
 repeat_region /note="LTR33 repeat: matches 6. .216 of consensus"
 12540. .12716
 repeat_region /note="MIR repeat: matches 20. .201 of consensus"
 12732. .13094
 repeat_region /note="L2 repeat: matches 1129. .1502 of consensus"
 13311. .13725
 repeat_region /note="L2 repeat: matches 1864. .2295 of consensus"
 13964. .14156
 repeat_region /note="MIR repeat: matches 58. .261 of consensus"
 14257. .14330
 repeat_region /note="MIR repeat: matches 79. .152 of consensus"
 14386. .14524
 repeat_region /note="MIR repeat: matches 47. .183 of consensus"
 15029. .15467
 repeat_region /note="Charlie4a repeat: matches 25. .507 of consensus"
 16483. .17040
 repeat_region /note="L1P4 repeat: matches 5585. .6142 of consensus"
 17775. .18439
 repeat_region /note="L1MB2 repeat: matches 5677. .6341 of consensus"
 18432. .18880
 repeat_region /note="L1M4 repeat: matches 4317. .4766 of consensus"
 18881. .19351
 repeat_region /note="L1P48 repeat: matches 5682. .6163 of consensus"
 19352. .19362
 repeat_region /note="L1M4 repeat: matches 4307. .4317 of consensus"
 19465. .20389
 repeat_region /note="L1M4 repeat: matches 3330. .4263 of consensus"
 20397. .20452
 repeat_region /note="28 copies 2 mer ta 87% conserved"
 20481. .21323
 repeat_region /note="L1MB5 repeat: matches 5313. .6138 of consensus"
 21799. .22116
 repeat_region /note="L1MB2 repeat: matches 5844. .6160 of consensus"
 22130. .22456
 repeat_region /note="MSTA repeat: matches 55. .426 of consensus"
 22457. .22511
 repeat_region /note="MLT2B repeat: matches 390. .444 of consensus"
 22512. .22535
 repeat_region /note="12 copies 2 mer ta 100% conserved"
 22615. .22989
 repeat_region /note="MLT2B repeat: matches 1. .382 of consensus"
 22990. .23047
 repeat_region /note="MSTA repeat: matches 1. .60 of consensus"
 23048. .23953
 repeat_region /note="L1MB6 repeat: matches 4929. .5842 of consensus"
 23954. .24953
 repeat_region /note="MER11C repeat: matches 1. .1071 of consensus"
 24955. .25128
 repeat_region /note="L1MB6 repeat: matches 4759. .4933 of consensus"
 25155. .25280
 repeat_region /note="L1 repeat: matches 4020. .4145 of consensus"
 25280. .25431
 repeat_region /note="L1 repeat: matches 5175. .5334 of consensus"
 25592. .26020
 repeat_region /note="L1 repeat: matches 4767. .5200 of consensus"
 26031. .26260
 repeat_region /note="L1M4A repeat: matches 5987. .6274 of consensus"
 26261. .26667
 repeat_region /note="MSTB repeat: matches 2. .422 of consensus"
 26673. .27291
 repeat_region /note="L1M4A repeat: matches 5318. .5988 of consensus"
 27284. .28932
 repeat_region /note="L1 repeat: matches 3281. .4957 of consensus"
 28933. .28976
 repeat_region /note="L1PAL3 repeat: matches 6113. .6156 of consensus"
 28985. .33897
 repeat_region /note="L1PAL3 repeat: matches 480. .5401 of consensus"
 34102. .34982
 repeat_region /note="L1PAL3 repeat: matches -651. .215 of consensus"

```

repeat_region 34990..35688
/note="L1M4c repeat: matches 1204..1905 of consensus"
repeat_region 35689..35968
/note="L1P4A repeat: matches 5867..6146 of consensus"
repeat_region 35969..35982
/note="L1M4c repeat: matches 1191..1204 of consensus"
repeat_region 35983..36287
/note="AluY repeat: matches 1..305 of consensus"
repeat_region 36288..36533
/note="L1M4c repeat: matches 959..1191 of consensus"
repeat_region 37203..37867
/note="L1MEC repeat: matches 1279..1885 of consensus"
repeat_region 37869..43115
/note="L1P4A repeat: matches 900..6146 of consensus"
repeat_region 43111..43891
/note="L1P4A repeat: matches 1..776 of consensus"
misc_feature 43445..43871
/note="CpG island"
/note="evidence=not experimental"
repeat_region 43894..44069
/note="L1MD repeat: matches 973..1193 of consensus"
repeat_region 44684..44856
/note="L1MD repeat: matches 0..174 of consensus"
repeat_region 45662..46050
/note="M3TA repeat: matches 1..426 of consensus"
repeat_region 46643..46875
/note="MIR repeat: matches 12..257 of consensus"
repeat_region 47951..47992
/note="21 copies 2 mer ac 90% conserved"
repeat_region 48970..49027
/note="29 copies 2 mer tt 70% conserved"
repeat_region 49033..49491
/note="L1P4A repeat: matches 5684..6146 of consensus"
repeat_region 52768..53073
/note="AluY repeat: matches 1..306 of consensus"
repeat_region 54837..55052
/note="MIR repeat: matches 25..258 of consensus"
repeat_region 56901..58716
/note="L1M4c repeat: matches 251..1634 of consensus"
repeat_region 58837..59017
/note="L1M4c repeat: matches 1872..2047 of consensus"
repeat_region 59019..59988
/note="L1MEC repeat: matches 2389..2986 of consensus"
repeat_region 59989..60713
/note="L1M4A7 repeat: matches 5553..6284 of consensus"
repeat_region 60714..60929
/note="L1MEC repeat: matches 2986..3204 of consensus"
repeat_region 60930..61183
/note="AluJo repeat: matches 1..299 of consensus"
repeat_region 61185..61550
/note="L1 repeat: matches 3187..3567 of consensus"
repeat_region 61558..61792
/note="L1P repeat: matches 3047..3292 of consensus"
repeat_region 61875..63977
/note="L1P repeat: matches 978..3085 of consensus"
repeat_region 63978..64287
/note="AluSg repeat: matches 1..310 of consensus"
repeat_region 64288..64668
/note="L1P repeat: matches 599..978 of consensus"
repeat_region 64787..64852
/note="L1 repeat: matches 3688..3750 of consensus"

Query Match 45.3%; Score 851.4; DB 9; Length 104534;
Best Local Similarity 99.9%; Pred. No. 1.2e-231;
Matches 852; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 730 CCAAGATCTGGCATTCCTTGCTCATAATGAGATAGTCTCCAAATTAGCCAAGATG 789
DB 11196 CCACAGATCTGGCATTCCTTGCTCATAATGAGATAGTCTCCAAATTAGCCAAGATG 11137

QY 790 AACCTCATGGACATCAACCAAGATCTTCTCCCTCGACGCCGACAAAGGAGGAGGAC 849
DB 11136 AACCTCATGGACATCAACCAAGATCTTCTCCCTCGACGCCGACAAAGGAGGAGGAC 11077

```

```

QY 850 ACTGACACAGAGGAGAGCAGGCTCTCAATCAAGCAGTGTATGACAACGACTCCTATACT 909
DB 11076 ACTGACACAGAGGAGAGCAGGCTCTCAATCAAGCAGTGTATGACAACGACTCCTATACT 11017

QY 910 TTGGACACAGCTTTTGGCCAGGAGCGTTACAAACGTTTTCATCAACAGCAGGAGTGGCTGG 969
DB 11016 TTGGACACAGCTTTTGGCCAGGAGCGTTACAAACGTTTTCATCAACAGCAGGAGTGGCTGG 10957

QY 970 GGTGTTCTTGGACACACCTTTGGCTTGGCTTCTTATGGCCACTTGAAGCTGTTTGGCAA 1029
DB 10956 GGTGTTCTTGGACACACCTTTGGCTTGGCTTCTTATGGCCACTTGAAGCTGTTTGGCAA 10897

QY 1030 GTCCCTCTTAGCCCATGCTGCTGATGTTGACGCTTGGATGTCAAGGACAGAGCCACAT 1089
DB 10896 GTCCCTCTTAGCCCATGCTGCTGATGTTGACGCTTGGATGTCAAGGACAGAGCCACAT 10837

QY 1090 TTCACTGTCTGTCAGTCAATGGCCATCTGGACTGTGTACGTGTCTTTTGAAGCTGGTGCC 1149
DB 10836 TTCACTGTCTGTCAGTCAATGGCCATCTGGACTGTGTACGTGTCTTTTGAAGCTGGTGCC 10777

QY 1150 TCTCTGTGTGTAGCATCTTACAACTGTTCTCCGCTGCTCACAGCTGCCGCTGATGGT 1209
DB 10776 TCTCTGTGTGTAGCATCTTACAACTGTTCTCCGCTGCTCACAGCTGCCGCTGATGGT 10717

QY 1210 GCTGTGTGTATCTCGCAGGAGCTCTAGACCATGTTGACAGGCGCAACGCTCAAGCTAAA 1269
DB 10716 GCTGTGTGTATCTCGCAGGAGCTCTAGACCATGTTGACAGGCGCAACGCTCAAGCTAAA 10657

QY 1270 CTACAGTCTGGGCATCAACATAGCTTTCATGTTCTGGCCCCCTCTATTGCGCCGAGTC 1329
DB 10656 CTACAGTCTGGGCATCAACATAGCTTTCATGTTCTGGCCCCCTCTATTGCGCCGAGTC 10597

QY 1330 TACGGGCACTGGACTGTTTCCGCTGCTTTCGCTCCAGCGGGCAGACCTGACTACAAAC 1389
DB 10596 TACGGGCACTGGACTGTTTCCGCTGCTTTCGCTCCAGCGGGCAGACCTGACTACAAAC 10537

QY 1390 TGCACTGACCGGGCTATTGGCTGCTGTCCTCAAGACCCCGCACCTCTCTTCAAAATCTGC 1449
DB 10536 TGCACTGACCGGGCTATTGGCTGCTGTCCTCAAGACCCCGCACCTCTCTTCAAAATCTGC 10477

QY 1450 CTCCATCATTAATTGTAGCCAGAGTATATCCAGCTGTTTAATCGATTTTGGTGTCTAATATC 1509
DB 10476 CTCCATCATTAATTGTAGCCAGAGTATATCCAGCTGTTTAATCGATTTTGGTGTCTAATATC 10417

QY 1510 TACCTTCCATCTCTCTCCCTTGACCTGACCTCAAGATGATAAAGGCATTGCAATGCTG 1569
DB 10416 TACCTTCCATCTCTCTCCCTTGACCTGACCTCAAGATGATAAAGGCATTGCAATGCTG 10357

QY 1570 CTACAGGCCCGAG 1582
DB 10356 CTACAGGCCCGAG 10344

RESULT 6
AF403040 1262 bp mRNA linear ROD 02-JAN-2002
LOCUS Mus musculus ankyrin repeat domain-containing SOCS box protein
DEFINITION Asb-12 mRNA, complete cds.
ACCESSION AF403040
VERSION AF403040.1 GI:18034099
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1262)
AUTHORS Kile, B.T. and Nicola, N.A.
TITLE SOCS box proteins
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1262)
AUTHORS Kile, B.T. and Nicola, N.A.
TITLE Direct Submission

```


JOURNAL	Submitted (27-JUL-2001) Cancer & Hematology, The Walter & Eliza Hall Institute of Medical Research, Royal Melbourne Hospital, Parkville, Victoria 3050, Australia		
FEATURES	Location/Qualifiers		
source	1. .1262		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/db_xref="taxon:10090"		
CDS	198..1124		
	/codon_start=1		
	/product="ankyrin repeat domain-containing SOCS box protein Asb-12"		
	/protein_id="AA157359.1"		
	/db_xref="GI:18034100"		
	/translation="NNLMIDIAKIFSLQPEKEEDTGTGEKALNAQVYDNDSCITLDHLLHQRYKRFINSRSGWGPCTPLRAASYGHLNCKVLLHGHADVDLSLDVKAQPLP TAVSHGHEUCVWMLAGACPSGSIYNNCSFVLITASRDGAFALIGDLGHGHEANVKA KLPWASNIASCGMLYGAAYGHLDPCFLLLLYADPDYNCFTDGLLSRVPQPTLL EICLHNCEPEYIQLIDFGANIYLPSPVDPSTQDDKGIKLLQARATPRSLLSQTR LVIRSLCRANQSQATDQIDIPVLISYLKHQ"		
BASE COUNT	318 a	338 c	272 g 334 t
ORIGIN	Query Match 38.8%; Score 729.2; DB 10; Length 1262; Best Local Similarity 81.5%; Pred. No. 6.9e-197; Matches 883; Conservative 0; Mismatches 193; Indels 8; Gaps 3;		
QY	736	ATCTGCATTTCTCTGTGCTCATATGAG--AATAGTTCTCCAATTAGCCAAGATGAACC	793
DB	145	ATCTGGAATTTCTCTCTCTCTGCAATGAGATGCTATTTCGAATAAACCAAGAAATGAACC	204
QY	794	TCATGACATCACCAAGATCTTCTCCCTCTCTGAGCCGACAGGAGGAGGAGGACACTG	853
DB	205	TCATGATATGCCAAGATCTTCTCTCTGCACTGAAAGAGGAGGAGGAGGACACTG	264
QY	854	ACACAGGAGGAGGAGGCTCTCAATCAAGCAGTGTATGACACACGACTCTCTATACTTTGG	913
DB	265	GCACCGGGGAAACACAGGCTCTCAATCAAGCTGTATGACATGACTCTCTGTACCCCTGG	324
QY	914	ACCAGTTTGGCCAGGAGCGTTCAACAGTTTTCATCAACAGCAGGAGTGGCTGGGGTG	973
DB	325	ACCACCTTCTACACGAGGACGTTATAACGGTTTCATCAACAGCAGGAGTGGCTGGGGTA	384
QY	974	TTCTGGGACACCTTGGCTGGCTGCTTCTATGCGCACTTGTAGCTGTTTGCAGTCC	1033
DB	385	TACCTGGAACACCTTGGCTGGCTGCTTCTATGCTCACTTAAATTTGTGAAGGCTC	444
QY	1034	TCTTAGCCCATGGTGTGATGTTGACAGCTTGGATGTCAAGGCACAGACGCCACTTTTCA	1093
DB	445	TCTTGAACATGCTGCTGATGTTGATGCTTGGATGTCAAGGCACACACCACTTTTCA	504
QY	1094	CTGCTGTAGTCAATGCCACTGCTGACTGTGTAGCTGTGCTTTTGGAGCTGGTCTCTC	1153
DB	505	CTGCTGTAGCCACCGTCTCTGAGTGGCTGAGAATGCTTTTGAAGCTGGTGGCTGTC	564
QY	1154	CTGGTGTAGCACTACACACGTTCTCCCGTGTCTACAGCTGCCCGTGTGGTGGCTG	1213
DB	565	CTAGTGTAGCATCTACACAAATGCTCTCTCTGTTCTCACTGCTCACGTGATGGGGCT	624
QY	1214	TTGCTATCTCGAGGAGCTCTTAGACCATGGTGCAGAGGCCAACGTCAAAGCTTAAACTAC	1273
DB	625	TTGCCATCTTACAGGAGCTCTTAGGCGATGTCGCCGAGGCTAATGTCAAGCTAATCTAC	684
QY	1274	CAGTCTGGGCATCAACATAGCTTTCATGTTCTGGCCCTCTATTATTTGGCCGAGTCTACG	1333
DB	685	CAGTCTGGGCGTCAAAATATAGCTTTCATGTTCTGGCCCTCTATCTGGCTGAGTCTATG	744
QY	1334	GGCACTGGACTCTTTCCGCTCTTTTGTCTCCACGGGCGAGACCTGACTACACTGCA	1393
DB	745	GGCACTTGAATTTTCCGCTCTTTTGTCTCTATGGGGGAGATCTCTGATTACACTGCA	804
QY	1394	CTGACCAAGGCGCTATTGGCTGCTGTCCCAAGACCCGCGACCTCTCTTGAATCTGCGCTC	1453

DB	805	CTGACCAAGGCGCTTTTAAGTCGTGTTCACAGCCTTCGACACTCTCTTGAATCTGCTTC	864
QY	1454	ATCATAATTTGTGAGCCAGAGTATATCCAGCTGTGTTAATCGATTTCGTGCTAATATCTACC	1513
DB	865	ATCATAATTTGTGAGCCAGAGTATATCCAGCTGTGTTAATCGATTTCGTGCTAATATCTACC	924
QY	1514	TTCCATCTCTCTCCCTTGACCTGACCTCACAGATGATAAAGGCATTTGATTCGTGCTAC	1573
DB	925	TTCCATCTCTCTCCCTTGACCTGACCTCACAGATGATAAAGGCATTTGATTCGTGCTAC	984
QY	1574	AGCCCGAGCCACTCCAGGTCACCTTCTATACAGGTCGGTTAGTCGTCGCGAGGCT	1633
DB	985	AAGCCCGAGCCACTCCAGGTCACCTTCTATACAGGTCGGTTAGTCGTCGCGAGGCT	1044
QY	1634	TGTGCGAGGCTGGCGAGCCACAGGCATCAACAGCTGGATATTCCTCCCATGTTGATTA	1693
DB	1045	TCTGCGGGGACACAGTCAAGAGCCAGGACGCTGGATATTCCTCCCATGTTGATTA	1104
QY	1694	GCTACCTAAACACCAACTGTAATCTTCAGCTCTCCCGAGAACTT-----ATGATGCTT	1748
DB	1105	GCTACCTCAACATCAATGATAAGTTTCACTTCTCAGGAACCTAATGTCATGATACCT	1164
QY	1749	CCGAAAA--CCACTGGGACTCAGTAGTGGAGGACATTACAGCTCATCCACTTACCT	1807
DB	1165	CCAAAAATAAACCTGGTAACCCAGGTTACTTGAGTGAATAATACACCTCATAACTTCT	1224
QY	1808	GGAG 1811	
DB	1225	GGCG 1228	
RESULT 7	BC012399 Homo sapiens, Similar to hypothetical protein FLJ20126, clone		
LOCUS	MGC:8876 IMAGE:3864128, mRNA, complete cds.		
DEFINITION	MGC:8876 IMAGE:3864128, mRNA, complete cds.		
ACCESSION	BC012399		
VERSION	BC012399.1 GI:15214550		
KEYWORDS	MGC.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 2653)		
AUTHORS	Strausberg, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-AUG-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk Email: cgabbs@mail.nih.gov Tissue Procurement: DCTD/DTF cDNA Library Preparation: Life Technologies, Inc. DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ Contact: villalon@bcm.tmc.edu Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.		
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAK Plate: 21 Row: j Column: 23 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7020020. Location/Qualifiers 1. .2653 /organism="Homo sapiens"		

```
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:8876 IMAGE:3864128"
/tissue_type="Ovary, adenocarcinoma"
/clone_lib="NIH_MGC_66"
/lab_host="DH10B"
/notes="vector: pcwv-sport6"
69..2183
CDS
  /codon_start=1
  /product="Similar to hypothetical protein FLJ20126"
  /protein_id="AAH12399.1"
  /db_xref="GI:15214551"
  /translation="MDHITVPKVENKLVDRYVSKPANGILYLTHLIYVEASGAA
    RKTWIALHIAIVTEKLPITSLGCLPLTRCKNFRVAHFVLDLVCHEVYIISLLKLSQ
    PALPDLVAFSYNPKSKEMRSKWKLDIPISDFGRMGI PNRNWTITDARNVEICST
    YPEEIVVPKSVTLGTWVSGSKFRKVRPVLSLYLKENNAACIRCQPSGLSPYTRCVD
    DELLLEAISOTNPGSQFMVVDTRPKLNAMANAAGKGVENEDNYANIRFRPMGIENI
    HVMRSISLOKLEVCCLKPTMSEFSLGSSGLWRHKAIMDAGIFITIAVKVEKASV
    LVHCSQDMORTAQVCSVASIILLDPFRTFKGLMLLEKEMWSMGHFKFSQRCHLGDGS
    KEVSPITQDLDCIQWLEQFPKAFNFENFLEIHDHVFPSQFQFNLCQKQREDL
    RVYEKTHSVMPFLVQRKPPFRNPLKGYFTMYGVLPSTPYPIQFCWGMNRFDKGLQ
    PKOSMLELLEIKKRAMLETVDHELEKLVKRDPEPSEICTCSQLGNILSHLGSPL
    TNPGLGFWINGDNLTMENGLTREGGLRAQMDQVKSOGADLHNNCEIIVGSIRAINI
    SGVDGISEAMGISDGMTPEATGFSKDLGICGAMDISEATGISGNLGISEARFSDGM
    GILGDTGISKATKEADYSKHQ"
BASE COUNT 749 a 566 c 645 g 693 t
ORIGIN
  Query Match 37.8%; Score 710.2; DB 9; Length 2653;
  Best Local Similarity 99.6%; Pred. No. 2e-191, 3; Indels 0; Gaps 0;
  Matches 712; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
  QY 23 CGAGGGTAGAAAACGTAATGGTGGATCGTTATGTGTAGTAAAGAACCCAGCTAATGGGA 82
  DB 88 CCAAGGTAGAAAACGTAATGGTGGATCGTTATGTGTAGTAAAGAACCCAGCTAATGGGA 147
  QY 83 TTCTTTTACTTACTGCAACCCACTGATCTATGTGGAGGCTTCAGGTGAGCCCGGAAAG 142
  DB 148 TTCTTTTACTTACTGCAACCCACTGATCTATGTGGAGGCTTCAGGTGAGCCCGGAAAG 207
  QY 143 AAACATGGATTGACCTCCATCATGTCATGCCACTGTGGAGAGTTTACCATCATCAGCTCGG 202
  DB 208 AAACATGGATTGACCTCCATCATGTCATGCCACTGTGGAGAGTTTACCATCATCAGCTCGG 267
  QY 203 GTTGTCCTCCCTGACCCCTCCGCTGCAAGAAATTCGGGTGGGCCCACTTTGTTTAGATTCTG 262
  DB 268 GTTGTCCTCCCTGACCCCTCCGCTGCAAGAAATTCGGGTGGGCCCACTTTGTTTAGATTCTG 327
  QY 263 ACCTTGTCGCATGAGGTTTATATTTCACTGCTCAAGCTTTCTAGCCAGCAATTACCTG 322
  DB 328 ACCTTGTCGCATGAGGTTTATATTTCACTGCTCAAGCTTTCTAGCCAGCAATTACCTG 387
  QY 323 AAGATCTTTATGCTTTTCTTATAATCCCAATTCCTCAAAAGAGATGAGGGAAGTGAT 382
  DB 388 AAGATCTTTATGCTTTTCTTATAATCCCAATTCCTCAAAAGAGATGAGGGAAGTGAT 447
  QY 383 GGAACCTGATGACCCCAATATCAGACTTTGGGGGATAGGGAATACCCACAGAAACTGGA 442
  DB 448 GGAACCTGATGACCCCAATATCAGACTTTGGGGGATAGGGAATACCCACAGAAACTGGA 507
  QY 443 CCATAACAGATGCCAACAGAACTATGAGATATGAGACCTTACCCTCTCGTAAATAGTGG 502
  DB 508 CCATAACAGATGCCAACAGAACTATGAGATATGAGACCTTACCCTCTCGTAAATAGTGG 567
  QY 503 TTCTCTAAATCTGTTTACCTTTGGGAACCGTGTGGAGTTTCAAGTTTCAAGATGAAGAAC 562
  DB 568 TTCTCTAAATCTGTTTACCTTTGGGAACCGTGTGGAGTTTCAAGTTTCAAGATGAAGAAC 627
  QY 563 GTGTCCTCTGCTCTCCTACCTCTACAAAGAGAAACAAATGCTGCAATTTGCCCGCTGTAGCC 622
  DB 628 GTGTCCTCTGCTCTCCTACCTCTACAAAGAGAAACAAATGCTGCAATTTGCCCGCTGTAGCC 687
  QY 623 AGCCTCTCTCGGATTTTACATCGCTGTGTAGATGATGAGCTCTTGTGGAGGCCATTA 682
```

```
Db 688 AGCCTCTCTCGGATTTTACATCGCTGTGTAGATGATGAGCTCTTGTGGAGGCCATTA 747
QY 683 GCCAAACAACCCAGGAGCCAGTTTATGTATGTCTAGACACAAAGCAAGAT 737
Db 748 GCCAAACAACCCAGGAGCCAGTTTATGTATGTCTAGACACAAAGCAAGAT 802

RESULT 8
AX406008 1345 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 423 from Patent WO0222660.
ACCESSION AX406008
VERSION AX406008.1 GI:21439455
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Tang, Y.T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q.A., Ren, F.,
Xue, A.J., Yang, Y., Wehrman, T. and Drmanac, R.T.
TITLE Novel nucleic acids and polypeptides
JOURNAL Patent: WO 0222660-A 423 21-MAR-2002;
HYSBQ, INC. (US)
FEATURES
  Location/Qualifiers
    1..1345
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    92..1204
    /note="unnamed protein product"
    /codon_start=1
    /protein_id="GI:21439456"
    /db_xref="GI:21439456"
    translation="MDHITVPKVENKLVDRYVSKPANGILYLTHLIYVEASGAA
      RKTWIALHIAIVTEKLPITSLGCLPLTRCKNFRVAHFVLDLVCHEVYIISLLKLSQ
      PALPDLVAFSYNPKSKEMRSKWKLDIPISDFGRMGI PNRNWTITDARNVEICST
      YPEEIVVPKSVTLGTWVSGSKFRKVRPVLSLYLKENNAACIRCQPSGLSPYTRCVD
      DELLLEAISOTNPGSQFMVVDTRPKLVKASVLVHCSQDMORTAQVCSVASIILLD
      PPFRTFKGLMLLEKEMWSMGHFKFSQRCHLGDGSKEVSPITQDLDCIQWLEQFPK
      AFENFLEIHDHVFSCQFNLCQKQREDLR"
BASE COUNT 380 a 300 c 321 g 344 t
ORIGIN
  Query Match 37.5%; Score 703.4; DB 6; Length 1345;
  Best Local Similarity 99.2%; Pred. No. 1.7e-189;
  Matches 707; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
  QY 23 CGAGGGTAGAAAACGTAATGGTGGATCGTTATGTGTAGTAAAGAACCCAGCTAATGGGA 82
  DB 111 CCAAGGTAGAAAACGTAATGGTGGATCGTTATGTGTAGTAAAGAACCCAGCTAATGGGA 170
  QY 83 TTCTTTTACTTACTGCAACCCACTGATCTATGTGGAGGCTTCAGGTGAGCCCGGAAAG 142
  DB 171 TTCTTTTACTTACTGCAACCCACTGATCTATGTGGAGGCTTCAGGTGAGCCCGGAAAG 230
  QY 143 AAACATGGATTGCACTCCATCATGTCATGCCACTGTGGAGAGTTTACCATCATCAGCTCGG 202
  DB 231 AAACATGGATTGCACTCCATCATGTCATGCCACTGTGGAGAGTTTACCATCATCAGCTCGG 290
  QY 203 GTTGTCCTCCCTGACCCCTCCGCTGCAAGAAATTCGGGTGGGCCCACTTTGTTTAGATTCTG 262
  DB 291 GTTGTCCTCCCTGACCCCTCCGCTGCAAGAAATTCGGGTGGGCCCACTTTGTTTAGATTCTG 350
  QY 263 ACCTTGTCGCATGAGGTTTATATTTCACTGCTCAAGCTTTCTCAGCCAGCAATTACCTG 322
  DB 351 ACCTTGTCGCATGAGGTTTATATTTCACTGCTCAAGCTTTCTCAGCCAGCAATTACCTG 410
  QY 323 AAGATCTTTATGCTTTTCTTATAATCCCAATTCCTCAAAAGAGATGAGGGAAGTGAT 382
  DB 411 AAGATCTTTATGCTTTTCTTATAATCCCAATTCCTCAAAAGAGATGAGGGAAGTGAT 470
```


Matches 709; Conservative 0; Mismatches 140; Indels 3; Gaps 1;			
QY	734	AGATCTGGCATTTCTTGTGCTCATATAGAGATAGTTC--TCCAAATAGCAAGATGA	790
Db	152248	AGATCGGAAATTTCTTCTTCTTACATAGAGATAGTCTATTCNATTAACAGATGA	152307
QY	791	ACCTCATGACATCAACCAAGATCTTCTCCCTCTGAGCCGACCAAGGAGGAGGACCA	850
Db	152308	ACCTCATGATATCGCAAGATCTTCTCTCTTCTGCAACAGAAAAAGAGGAGGAGGACG	152367
QY	851	CTGACACAGAGAGAGAGGCTCTCAATCAAGCAGTGTATCAACAGCTCTTATCTT	910
Db	152368	CCGACACAGCGGAAAGAGGCTCTCAATCAAGCTGTATATGACATGACTCTTGTACCC	152427
QY	911	TGACACAGCTTTTGCCGACAGGAGCGTTACAAAGTTTTCATCAACAGCAGGAGTGGCTGG	970
Db	152428	TGACCACTTCTACACAGGAGGTTTAAAGGTTTTCATCAACAGCAGGAGTGGCTGG	152487
QY	971	GTGTTCTCTGGACACCCCTTGGCTTGGCTGCTTCTTATGGCCACTTTGAGCTGT	1030
Db	152488	GTATACCTTGGACACCCCTTGGCTTGGCAGCTCTTATGTCACCTTAGATTGTGGAAG	152547
QY	1031	TGCTCTTACCCATGTGCTGATGTTGACAGCTTGGATGTCAAGGACAGACGCACTTT	1090
Db	152548	TGCTCTGGAACATGCTGCTGATGTTGATAGCTTGGATGTCAAGGACCAACACCACTTT	152607
QY	1091	TCACTGCTGTCAGTCATGCGCATCTGCACTGTGCTGCTGCTTTTGGAGCTGTGCTCT	1150
Db	152608	TCACTGCTGTTAGTCAGGTCATCTGAGTGTGTGAGAACTTTTGAAGCTGTGCTCT	152667
QY	1151	CTCTGCTGCTAGCATCTCAACAACTGTTTCTCCGCTGCTCAGCTGCGCCGATGGTG	1210
Db	152668	GTCTAGTGTGAGCATCTCAACAACTGTTTCTCTCTGCTCACTGCTCACTGATGGG	152727
QY	1211	CTGTTGCTATCTGAGAGCTCTTAGACATGCTGTCAGCTGCGCCGATGGTGATG	1270
Db	152728	CCTTTGCCATCTTACAGGAGCTCTAGGCGATGCTGAGAGGCTATGTCAAGGCGAAG	152787
QY	1271	TACCACTGCGGCATCAACATAGCTTCATGTTCTGCGCCCTCTATTTGGCGGAGTCT	1330
Db	152788	TGCAGCTGCGGCATCAACATAGCTTCATGTTCTGCGCCCTCTATTTGGCTGCACT	152847
QY	1331	ACGGGACCTGCACTGCTTTCCGCTGCTCTTTGCTCCACGCGGCGAGACCTGATCAACT	1390
Db	152848	ATGGTATCTTGATGTTTCCGCTGCTTTGCTCTATGGCGAGATCTGATTACACT	152907
QY	1391	GCACTGACAGGCGCTTATGGCTGCTGTCGCAAGACCCGCGACCTCTCTGAAATCTGCC	1450
Db	152908	GTATTGACAGGCGCTTAAAGTCGAGTTCACAGGCTCGCACACTCTCTGAAATCTGCC	152967
QY	1451	TCATCATATTTGTGAGCCAGATATATCCAGCTGTTAATCGATTTTGTGCTATATCT	1510
Db	152968	TTATCATATTTGTGAGCCAGATATATCCAGCTGTTAATAGATTTTGTGAGCGAATCT	153027
QY	1511	ACCTTCCATCTCTCTCCCTTGACCTGACCTCAAGATGATAAGGCACTTGCTTCTGCG	1570
Db	153028	ACCTCCATCTCTACCTTTGGACCCCAACATCTCAAGATGATAAGGCACTAGTCTGCTG	153087
QY	1571	TACAGGCCCGAG 1582	
Db	153088	TACAAGCCCGAG 153099	
RESULT 11			
AC108644			
LOCUS			
DEFINITION Rattus norvegicus clone CH230-59A6, *** SEQUENCING IN PROGRESS ***			
AC108644			
ACCESSION			
AC108644.5 GI:24818595			
VERSION HTG; HTGS_P1; HTGS_DRAFT; HTGS_ENRICHED.			
KEYWORDS Rattus norvegicus (Norway rat)			
SOURCE Rattus norvegicus			
ORGANISM			

REFERENCE
AUTHORS

1 (bases 1 to 225982)

Muzny,D., Marie, Metsker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biwalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., De Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.B., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshwar, L., Louisedge, H., Lozano, R.J., Lu, X., Ma, J., Maheshwar, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Manthey, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankweli, O., Okwionu, G., Olarnpungoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pu, L., Pu, L., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wleczyk, R., Woodden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.

TITLE
JOURNAL

2 (bases 1 to 225982)

Unpublished

Worley, K.C.

Direct Submission

Submitted (31-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 225982)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (09-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 9, 2002 this sequence version replaced gi:23267762. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GPSJ

Center clone name: CH230-59A6

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 211822 bases at least Q40

Consensus quality: 215372 bases at least Q30

Consensus quality: 217787 bases at least Q20

Estimated insert size: 216316; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

- * NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
- * NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
- * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

- * 1 32687: contig of 32687 bp in length
- * 32688 32787: gap of unknown length
- * 32788 102606: contig of 69819 bp in length
- * 102606 102706: gap of unknown length
- * 102706 123303: contig of 20597 bp in length
- * 123303 123403: gap of unknown length
- * 123403 123404: contig of 88906 bp in length
- * 123404 212310: gap of unknown length
- * 212310 212409: gap of unknown length
- * 212409 213674: contig of 1264 bp in length
- * 213674 213774: gap of unknown length
- * 213774 214949: contig of 1176 bp in length
- * 214949 215049: gap of unknown length
- * 215049 217109: contig of 2060 bp in length
- * 217109 217210: gap of unknown length
- * 217210 225982: contig of 8773 bp in length.

FEATURES

source

```
1. .225982
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-59A6"
complement(4374..5073)
/note="clone_boundary"
clone_end:17
site:EcORI
```

misc_feature

```
end_sequence:BH331578
32788..35656
/note="wgs_contig"
```

misc_feature

```
81431..82225
/note="clone_boundary"
clone_end:Sp6
site:EcORI
end_sequence:BH331579"
```

BASE COUNT 65353 a 44744 c 44903 g 64320 t 6662 others
ORIGIN

Query Match

Best Local Similarity 32.6%; Score 612; DB 2; Length 225982;

Matches 709; Conservative 0; Mismatches 140; Indels 3; Gaps 1;

```
QY 734 AGATCTGGCATTTCTCTGTGCTCATATGAGAAATAGTTC---TCCAATTAGCCAAGATGA 790
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130053 AGATCGGGAATTTCTCTCTTTTACAAATGAGAAATAGTTCATATCCATTAACCAAGATGA 130112
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 791 ACCTCATGACATACCAAGATCTTCTCCTCTCTGAGCCGCAACAGGAGGAGGAGACA 850
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130113 ACCTCATGATATCGCAAGATCTTCTCTCTCTGCAACAGAAAGAGGAGGAGGAGCG 130172
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 851 CTGACACAGAGAGAGAGGAGGCTCTCAATCAAGCAGTGTATCAACAGCTCTCTATACCT 910
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130173 CCGACACAGGCGAAAGAGGCTCTCAATCAAGCTGTATATGACATGACTCTTTGTACCC 130232
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 911 TCGACCAAGCTTTTGGCGCAGGAGCGTTACAAACGTTTTCATCAACAGCAGAGAGTGGCTGG 970
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130233 TCGACCAAGCTTTTACACAGGAGCGTTATAACGGTTTCATCAACAGCAGAGAGTGGCTGG 130292
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 971 GTGTTCTCTGGGACACCCCTTGGCGCTTGGCTGTCTTATATGGCCACTTGAAGCTTTTGC 1030
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130293 GTATACCTGGAACACCCCTTGGCGCTTGGCAGCCTCTTATGGTCACCTTAGATTGTG 130352
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1031 TCCTCTTAGCCCATCGTCTGATGTGTAGCAGCTTGGATGTCAAGGACAGACGCCACTTT 1090
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130353 TCCTCTGGAACATGCTGATGTGTAGCTTGGATGTCAAGGACAGACGCCACTTT 130412
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1091 TCACCTGCTGTAGTCAATGGCCATCTGGACTGTGTAGTGTCTTTTGGAGCTTGGTGCCT 1150
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130413 TCACCTGCTGTAGTCAATGGCTCATCTGGAGTGTGTAGATACTTTTAGAAGCTTGGCT 130472
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1151 CTCCTGGTGTAGCATCTACAAACTGTTTCCCGTGTCTACAGCTGCCCGTGTGGTG 1210
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130473 GTCTAGTGGTAGCATCTACAAACTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 130532
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1211 CTGTTGCTATCTCTGAGGAGCTCTAGACCATGGTGTGAGAGGCGCAACGTCAAAGCTAA 1270
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130533 CTTTGGCCATCTTACAGGAGCTCTTAGGCGATGGTGTGAGAGGCTAATGTCAAGGCGAA 130592
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1271 TACCACTGTGGGCATCAAAACATAGCTTCTATGTTCTGGGCCCCCTCTATTTGGCGGAGTCT 1330
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130593 TCCGAGTCTGGGCATCAAAACATAGCTTCTATGTTCTGGGCCCCCTCTATTTGGCTG 130652
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1331 ACGGCACTGACATGTTTTCGCGCTGCTTCTCCAGGGGCGACACCTGACTACAAC 1390
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130653 ATGGTCATCTTGATTGTTTTCGCGCTGCTTCTCTATGAGGCGAGATCTCTGATTACA 130712
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1391 GCACCTGACAGGCGCTATTTGGCTGTGTCCCAAGACCCGCGACCCCTCTTTGAAATCTGC 1450
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130713 GTATTGACCGAGCCCTCTTAAGTTCGAGTTCCACAGCCTCGCACCTCTCTTGAATCTGC 130772
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1451 TCCATCAATTTGTGAGCGCAGAGTATATCCAGCTGTTAATCGATTTTGTGTCTAATATCT 1510
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130773 TCCATCAATTTGTGAGCGCAGAGTATATACAGCTTTTATAGATTTTGGAGCGAATCT 130832
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1511 ACCTTCCATCTCTCTCTCTTGAACCTGACCTCAAGATGATAAAGGCAATTCATTGCTGC 1570
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130833 ACCTTCCATCTCTCTCTTGGACCCCAACATCTCAAGATGATAAAGGCAATTCATTGCTGC 130892
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1571 TACAGGCGCGAG 1582
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130893 TACAGGCGCGAG 130904
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 12

LOCUS

AL844570

DEFINITION

Zebrafish DNA sequence from clone DKEY-20D18, complete sequence.

ACCESSION

AL844570

VERSION

AL844570.9

KEYWORDS

HTG.

SOURCE

Danio rerio (zebrafish)

ORGANISM

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

REFERENCE AUTHORS TITLE JOURNAL

1 (bases 1 to 240816)
Sehra, H.
Direct Submission
Submitted (11-FEB-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 10, 2003 this sequence version replaced gi:27652710.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zfish-help@sanger.ac.uk

COMMENT

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep Repeat names
beginning 'Dr' were identified by The Recon repeat discovery system
(Zhiron Bao and Sean Eddy, submitted), and those beginning 'drr'
were identified by Rick Waterman (Stephen Johnson lab, WashU). For
further information see http://www/Projects/D_rerio/fishmask.shtml
DKEY-20D18 is from a Zebrafish BAC library
VECTOR: pIndigoBAC-5.

FEATURES

source 1..240816
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-20D18"
/clone_lib="DanioKey"
BASE COUNT 76749 a 43368 c 43554 g 77145 t
ORIGIN
Query Match 15.5%; Score 290.6; DB 5; Length 240816;
Best Local Similarity 61.0%; Pred. No. 3.9e-71;
Matches 489; Conservative 0; Mismatches 309; Indels 3; Gaps 1;
QY 782 CCAAGATGAACCTCATGGACATCACCAAGATCTTCTCCCTGCGACGCCGACAGGAGG 841
D 18732 CCAGCATGAGTTGATGGACATCTCTAAGATCTTTCTCTCTCTCAACCCAAAGAGGATG 18791
QY 842 AGGAGGACACTGCACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 901
D 18792 AGGAGGAA---GACACCAACTGCGAGTCAGCACTAATCAAGTGCTTTTCACAGACGCG 18848
QY 902 CTTATCTTTGGACCAAGCTTTTTCGCGCAGGAGGCTTACAAAGCTTTTCATCAACAGCAGGA 961
D 18849 ACAAACTCTTACTGAACCTGTTGCTTCAGGACAAATATAGAAGTCATCAACACCGCA 18908
QY 962 GTGGCTGGGGTGTCTTGGGACACCTTGGCTGTGCTCTTCTTATGCGCACTTGAGCT 1021
D 18909 GTGGTGGGGGATCCCACTCACTCCACTGCGACGCGGCTGTCTCAGGGTCACTACGCT 18968
QY 1022 GTTGAAGTCTCTTAGCCCACTGTGTGATGTTGACAGCTGGATGTCAGGACACAGA 1081
D 18969 GTCTGAGGTCTCTCTGGCTCAACGCGGCGAGAGGTCGACGCTGGATGTCAAAGCTCAAA 19028
QY 1082 CGCCACTTTTCTGCTGTCTGAGTCATGGCCATCTGGAATGTGTACGTGTCTTTTGAAG 1141

Db 19029 CACCACCTTTTACAGCCGCTCTCTGCGAAACACATAGATTGTGTGCTTTATTAAAGAG 19088
QY 1142 CTGGTGCTCTCTCTGGTGGTAGCATCTACACAACTGTTCCTCCCTGCTCAGAGTGCCC 1201
Db 19089 CTGGGCGAGACCTTAATGAAGCCCAACATAAATGCTCTCCCTGTGTGTGACTGGGGCA 19148
QY 1202 GTGATGGTGTCTTCTATCTCGAGGAGCTCTAGACCATCTAGACCATCTGTCAGAGGCAAGCTCA 1261
Db 19149 GAGAGGGGATGTGCACATCTGAAGAGCTCTTCAATATGCGCGCATGAGAGCTCA 19208
QY 1262 AAGCTAACTACAGCTGCGGCATCAACATAGCTTCATGTTCTGCGCCCTCTATTGG 1321
Db 19209 AGCCCCAAATGCGACACTGGGCTTTAAATGCGCACAGCTTGCCGAGGACCACTGTACATTT 19268
QY 1322 CGCAGCTCTAGGGCAGCTGGAAGCTCTTCCGCTCTCTTTCCTCCAGGGGAGACCCCTG 1381
Db 19269 CAGCGGTGTATGTCATCTAGGTTGTTTCAATGTTGCTGCGAGTGGAGCAACCCAG 19328
QY 1382 ACTACAACTGCACTGACAGGCGCTTATGGCTGCTGCTCCCAAGACCCCGACCTCTCTTG 1441
Db 19329 ACTACAACTGCACTGAGGAAAGATGCTGGCCAGGATCAAGAGGCTTAAACCTGTTTGG 19388
QY 1442 AAATCTGCTCATCAATATGTCAGCAGAGATATCCAGCTGTTTAATCGATTTTGGTG 1501
Db 19389 AGATGTGTCTGAGACATGCTGTGGCTGGAATACATTCAGCTGCTCATAGACTTTGGAG 19448
QY 1502 CTAATATCTACCTCCATCTCTCTCCCTGACCTGACCTCACAAGATGATAAGGATTTG 1561
Db 19449 CAAATGTTTATCTGCCCCCTTTAGTTGGAGATAAACCAACCACTGAGAGATGAAGCAGTAG 19508
QY 1562 CATTCTGCTCAGAGCCCGAG 1582
Db 19509 TCCTTCTGTTAAAGAGAGAG 19529

RESULT 13

BC053411 2354 bp mRNA linear VRT 09-JUN-2003
LOCUS Danio rerio cDNA clone MGC:56405 IMAGE:5605218, complete cds.
DEFINITION BC053411
ACCESSION BC053411.1 GI:31418217
VERSION MGC.
KEYWORDS
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 2354)
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hong, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahy, J., Helton, E., Kettner, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalhus, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL MEDLINE 22388257
PUBMED 12477932
REFERENCE 2 (bases 1 to 2354)


```
Db 239 GCTCAGGGAAGCGGAAAGGTTGAAATGTCGCGTTGGTAGATCGAGTGTCTCTCTAAAAA 298
Qy 70 CCAGCTAATGGGATCTCTTATCTTACTGCAACCCACCTGATCTATGTGGAGGCTTCAGT 129
Db 299 GCAGCTCTAGGTACTTGTATTTGCGGCTACCCATGTCTATATCTGGGAAATTCACCT 358
Qy 130 GCAGCCCGGAAAGAAACATGGATTGCATCTCCATCACAATGCCACTGTGGAGAAGTTTACCC 189
Db 359 GAGCGAAGAAAGAAACATGGATTCTTCACAGTCAGATTTCCACCATTCGAGAAACAGGCA 418
Qy 190 ATCACTAGCTGGTGTGCTCCCTGACCCCTCCGTCGAAGAAATTCGGGTGGCCCACTTT 249
Db 419 ACAACGCTACCGGATGCCCTCTGCTGATTCGTCGAAGAACTTTTCAGATAATACAGCTC 478
Qy 250 GTTTTAGATTCTGACCTGTGTGCGCATGAGTTTATATTTCACTGCTCAAGCTTTCTCAG 309
Db 479 ATCATACCTCAGGAAGAGATTGCCACGAGTGTATCTCTCCGTATACGCTTTCGAAG 538
Qy 310 CCAGCATTAACCTGAAGATCTTTATGCTTTTCTTATAATCCCAATCCCTCAAAAGAGATG 369
Db 539 CCAGTGAATATGAGGAGTTATACCTGTTTCAATCAACCCCATGCTGGATAAAGAGAA 598
Qy 370 AGGAAAGTGGATGGAACATGATTGACCAATATCAGACTTTTGGGGGTATGGGAATACCC 429
Db 599 AGAGCAAGGCTGGGTGCTGATCGATCTTAGTGAAGAAATACACGGGATGGGCTCCCT 658
Qy 430 AACAGAACTGGACCATACAGATGCCACAGAACTATGATATGACGACCTACCT 489
Db 659 AATCATTAATGCGCAGCTCAGCATGATGAAATAGAGACTATAGACTCTGTGACTCTTATCT 718
Qy 490 CTTGAATAGTGGTCTTAAATCTGTTACTCTGGGAAAGGTTGTTGGAAGTTCAAAAGTTC 549
Db 719 ACTGAATCTGCTTCCCAATCGGCCACGACATCATATAGTGGGAGTTCCAAATTC 778
Qy 550 AGAAGTAAGAACGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 609
Db 779 CGAGTAGAGCGGCAATTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 838
Qy 610 TCCGCTGTAGCAGCCTCTCTGATTTTACACTCGCTGCTGTATGATGATGATGATGATGAT 669
Db 839 TCCGCGAGCAGCAGCCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 898
Qy 670 TTGGAGCCATAGCCAAACCAACCCAGGAGCCAGTTTATGTTATGTTTATGATGATGATGAT 729
Db 899 CTCAGGCCATAGGAAGCAATCCAGGAAGTACTTGGTTATGCTGTGATGATGATGATGAT 958
Qy 730 CCAA 734
Db 959 CCTAA 963
```

RESULT 15

```
BC044359
LOCUS
DEFINITION
  Danio rerio, Similar to myotubularin related protein 6, clone
  MGC:55289 IMAGE:2600156, mRNA, complete cds.
ACCESSION
  BC044359
VERSION
  BC044359.1 GI:27882482
KEYWORDS
  MGC.
SOURCE
  Danio rerio (zebrafish)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
    Cypriniformes; Cyprinidae; Danio.
  1 (bases 1 to 2754)
  Strausberg, R.
  Direct Submission
  Submitted (17-JAN-2003) National Institutes of Health, Mammalian
  Gene Collection (MGC), Cancer Genomics Office, National Cancer
  Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
  USA
  NIH-MGC Project URL: http://mgc.nci.nih.gov
  Contact: MGC help desk
```

Email: cgabs-r@mail.nih.gov
Tissue Procurement: Dr. Sumio Sugano and Dr. Ko-ichi Kawakami
cDNA Library Preparation: Dr. Sumio Sugano and Dr. Ko-ichi Kawakami
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadan@systemsbio.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 101 Row: a Column: 20.

FEATURES

Location/Qualifiers
1..2754
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="MGC:55289 IMAGE:2600156"
/tissue_type="Whole body, adult, (one male and one female,
including unfertilized eggs)"
/clone_lib="Sugano_Kawakami_zebrafish_DRA"
/lab_host="DH10B"
/note="Vector: pME18S-FL3"
195..2126
/codon_start=1
/product="similar to myotubularin related protein 6"
/protein_id="AAH44359.1"
/db_xref="gi:27882483"
/translation="MEHRIKPEQVQLKDRFSNKSVTGTLTLTATHLIFVNNSTSS
QBIWTHHHIAPVEKLSLTTCPLLIQCRNFRVHVFIIPRERDCHDIYSSLLRLRP
VSYDELAFSYNPKNDOOREGEMOLIDAAEFERMGVPCDOMQLTDVNRDVKVCRTY
PRDLVPTASKPILVGSKRSKRPVPLTYFYQEKAAVCRSQPLSGFSARCLED
EHMLQAIKANHNSRNIIVMDTRPLNALANRAAGYENEDNINRIFQFVGENIH
VMRSLQKLELVGTSRISLTDYLVLENSGWLRIKAIKADAIKAKAVSVEGASVL
VHCSGMDRTAQVCAKLSLMDPYRTIKGFVLIKDWISFGHKFADRCQDLDDPK
EVSPTFTQFLECVWLTQFPOAFSEWFLIQIHEHISCFQGNFLGNSQRELDQ
LKERTYSLWHLKESNRELSTVQALTKLQNLGVSTPLKPTPPRERSLPBPRS
LILGAPLGRKDQDEBEMFAEAREDDPPALTGGERIVESSATENAYTELPGSFG
SRNEPAMVTLFEGVARMTC"

CDS

BASE COUNT 748 a 628 c 684 g 694 t
ORIGIN
Query Match 12.7%; Score 238.2; DB 5; Length 2754;
Best Local Similarity 59.3%; Pred. No. 2.4e-56;
Matches 424; Conservative 0; Mismatches 288; Indels 3; Gaps 1;
Qy 23 CGAGGTAGAAAACGTAATTTGGTGGATCGTTATGTGATGAAGAAACCCAGCTAAATGGA 82
Db 214 CAAAGGTAGAACAGTTAAGTCTGGATCGCT---TTAGCAACAAGTCCGTACAGGA 270
Qy 83 TTCTTTATCTTACTGCAACCCACCTGATCTATGTGGAGGCTTCAGGTGAGCCCGGAAAG 142
Db 271 CGCTGATCTGAGCGCCACACATTTGATCTTCTGGAGAATAATCAACAGCTCCACG 330
Qy 143 AAACATGATTCACCTCCATCAGTCCATGTCAGTGGAGAAGTTACCCATCAGTCTGG 202
Db 331 AGATATGGATCTCTGATCACCACATCGCCCCCGTGGAGAAGCTGTCCTCTCAACACCTG 390
Qy 203 GTTGTCCCTGACCTCCGCTGCAAGAAATTTCCGGGTGGCCCACTTTGTTTAGATTCTG 262
Db 391 GCTGCCACTCTCTCATTCAGTCCGAAATTTCCGAGTAGTCCACATTTGATTCGAAGAG 450
Qy 263 ACCTTGTGTGCCATGAGGTTTATATTTCACTGCTCTCAAGCTTTCTCAGCCAGCATTTACCTG 322
Db 451 AGCGGAGCTGCCACGACATCTACAGCTCTCTCTTAGGCTTCTGAGACCGGTGCTCTACG 510
Qy 323 AAGATCTTTATGCTTTTCTTTATATCCCAATCTCCAAAGAGATGAGGAAAGTGGAT 382
Db 511 ACGAGCTGTACGCTTTCTCTTACAAACCCAAAGCAGAACCGACGACGAGAGAGAGGCT 570
Qy 383 GGAAACTGATTGACCCCAATATCAGACTTTGGGGGTATGGGAATATCCCAACAGAACTGGA 442

Db	571	GGCAGCTCATAGACCTCGCGCTGAGTTTGGCGGATGGGTGCGCATGTGACCAAGTGC	630
Qy	443	CCATAACAGATGCCACAGAACTATGAGATATGCAGCACTACCCCTCCTGAAATAGTGG	502
Db	631	AGGTGACAGATGTTAACAGGGACTACAAAGGTGCGGAACATATCCAGAGACTTATACG	690
Qy	503	TTCTTAATCTGTACCTTGGGAACGGTGGTTGGAGTTCAAAGTTTCAGAAGTAAAGAAC	562
Db	691	TCCCCATTACAGCCAGTAAACCCATCATATGTCGGCAGCTCCAATTTCAGAAAGCAAGGCC	750
Qy	563	GTGTCCCTGTGCTCTCTACTCTTACAAAGAGAAACAATGCTGCCAATTTGCCGTGTAGCC	622
Db	751	GCTTTCTCTGCTGACGTATTTCTACCAGGAGAAAGGAGGAGTGTGTCGGGTAGTC	810
Qy	623	AGCCTCTCTCTGGATTTTACACTCGCTGTGTAGATGATGAGCTCTTGTGGAGGCCATTA	682
Db	811	AGCCTCTTTCTGGCTTCAGCGCTCGCTGTTTGAAGACGAGCACATGCTTCAGGCCATTA	870
Qy	683	GCCAAACAAACCCAGGAGGAGCCAGTTTATGTATGTTGTAGACACAGACCAAGAT	737
Db	871	GCAAGCCATCACACAGCCGAAATATATATGTGATGGACACGACCTAAGTT	925

Search completed: December 1, 2003, 12:30:45
Job time : 6742 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 1, 2003, 10:29:49 ; Search time 615 Seconds
(without alignments)
10060.858 Million cell updates/sec

Title: US-09-941-831a-6
Perfect score: 1878
Sequence: 1 catgattacgccagcttgg.....taaaaaa1878

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
Searched: 2190069 seqs, 1647345023 residues
Total number of hits satisfying chosen parameters: 4380138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1878	100.0	1878	10	US-09-941-831-6
2	1878	100.0	1878	12	US-10-207-175-13
3	1117.8	59.5	1222	10	US-09-908-193-13
4	931	49.6	1069	10	US-09-908-193-11
5	317	16.9	522	10	US-09-833-381-62
6	108.8	5.8	2019	10	US-09-908-805B-24
7	71.4	3.8	550	11	US-09-991-936-1472
8	71.2	3.8	389	10	US-09-867-701-5266
9	71.2	3.8	419	10	US-09-908-805B-26
10	69.4	3.7	401	9	US-09-922-217-28
11	69.4	3.7	401	10	US-09-833-263-28
12	69.4	3.7	401	13	US-10-025-380-28
13	67.2	3.6	3482	11	US-09-919-039-133
14	63.4	3.4	1675	10	US-09-833-381-1155
15	61.4	3.3	450	9	US-09-864-761-10608
16	60.2	3.2	2361	12	US-10-128-174-11

17	60.2	3.2	2370	14	US-10-164-080-1	Sequence 1, Appli
18	60.2	3.2	2370	14	US-10-299-327-1	Sequence 1, Appli
19	60.2	3.2	3516	11	US-09-866-050A-257	Sequence 257, App
20	60.2	3.2	3516	14	US-10-152-661-257	Sequence 257, App
21	55.4	2.9	6156	12	US-10-369-978-3	Sequence 3, Appli
22	55.4	2.9	24358	12	US-10-369-978-1	Sequence 1, Appli
23	53.6	2.9	449	12	US-09-814-353-1023	Sequence 1023, Ap
24	53.6	2.9	449	12	US-09-814-353-7392	Sequence 7392, Ap
25	53.6	2.9	521	12	US-09-814-353-13777	Sequence 13777, A
26	53.6	2.9	3811	14	US-10-177-293-105	Sequence 105, App
27	53.6	2.9	5352	10	US-09-833-381-1027	Sequence 1027, Ap
28	53.6	2.9	5352	10	US-09-833-381-1028	Sequence 1028, Ap
29	53	2.8	6549	12	US-10-301-822-92	Sequence 92, Appli
30	52.4	2.8	5148	12	US-10-021-571-1	Sequence 1, Appli
31	52	2.8	704	12	US-09-814-353-19659	Sequence 19659, A
32	52	2.8	1797	9	US-09-735-368-3	Sequence 3, Appli
33	51.8	2.8	1568	12	US-10-199-937-98	Sequence 98, Appli
34	51.8	2.8	2409	10	US-09-964-899-40	Sequence 40, Appli
35	51.8	2.8	3508	12	US-10-199-937-1	Sequence 1, Appli
36	51.8	2.8	4127	12	US-10-199-937-100	Sequence 100, App
37	51.8	2.8	4275	11	US-09-972-115A-5	Sequence 5, Appli
38	51.8	2.8	4406	12	US-10-199-937-106	Sequence 106, App
39	51.8	2.8	4992	12	US-10-199-937-134	Sequence 134, App
40	51.8	2.8	5002	12	US-10-199-937-132	Sequence 132, App
41	51.8	2.8	5778	14	US-10-163-587A-14	Sequence 14, Appli
42	51.8	2.8	6018	12	US-09-849-602-11	Sequence 11, Appli
43	51.6	2.7	4297	11	US-09-972-115A-3	Sequence 3, Appli
44	51.4	2.7	3984	12	US-10-199-937-3	Sequence 3, Appli
45	51.4	2.7	4134	9	US-09-841-835-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-941-831-6
; Sequence 6, Application US/09941831
; Patent No. US20020160493A1
; GENERAL INFORMATION:
; APPLICANT: Ebner et al.
; TITLE OF INVENTION: PTO49P1
; FILE REFERENCE: Serine/Threonine Phosphatase Polynucleotides, Polypeptides, and Antib
; CURRENT APPLICATION NUMBER: US/09/941,831
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: PCT/US01/06256
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/186,350
; PRIOR FILING DATE: 2000-03-02
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1878
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-941-831-6

Query Match 100.0%; Score 1878; DB 10; Length 1878;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CATGATTACGCCAAGCTTGGCAGAGGGTAGAAAAACGTGAAATTTGGTGGATCGTTATG	60
Db	1	CATGATTACGCCAAGCTTGGCAGAGGGTAGAAAAACGTGAAATTTGGTGGATCGTTATG	60
QY	61	AGTAAAGAACACGACGTAATGGGATTTCTTATCTTACTGCAACCCACCTGATCTATGTGGAG	120
Db	61	AGTAAAGAACACGACGTAATGGGATTTCTTATCTTACTGCAACCCACCTGATCTATGTGGAG	120
QY	121	GCTTCAGGTGACGCCGGAAGAAAACATGGATTGACATCCATCATTGCCACTGTGGAG	180
Db	121	GCTTCAGGTGACGCCGGAAGAAAACATGGATTGACATCCATCATTGCCACTGTGGAG	180
QY	181	AGTTTACCACATCAGTAGCTGGGTTGTCCCTGACCCCTCGCTGCAGAAATTTCCGGGTG	240

```
Db 181 |AGTTACCCATCACTAGCTGGGTGTGCCCTGACCTCGCTGCAAGAAATTTCCGGGTG 240
Qy 241 |GCCACTTTGTTTATGATCTGACCTGTGTGTCATGAGTTTATATTTTCACTGCTCAAG 300
Db 241 |GCCACTTTGTTTATGATCTGACCTGTGTGTCATGAGTTTATATTTTCACTGCTCAAG 300
Qy 301 |CTTTCTCAGCCAGCATTTACCTGAAGATCTTTATGCTTTTCTTTATATCCCAATCCTCA 360
Db 301 |CTTTCTCAGCCAGCATTTACCTGAAGATCTTTATGCTTTTCTTTATATCCCAATCCTCA 360
Qy 361 |AAAGAGATGAGGGAAGTGGATGGAACCTGATTTGACCCCAATATCAGACTTTGGGCGTATG 420
Db 361 |AAAGAGATGAGGGAAGTGGATGGAACCTGATTTGACCCCAATATCAGACTTTGGGCGTATG 420
Qy 421 |GGAATACCAACAGAACTGGACCAATACAGATGCCAAGCAAGAACTATGAGATATCGAGC 480
Db 421 |GGAATACCAACAGAACTGGACCAATACAGATGCCAAGCAAGAACTATGAGATATCGAGC 480
Qy 481 |ACCTACCTCCTGAAATAGTGGTTCCTAAATCTGTTACCTTGGGAACGGTGTGGAAGT 540
Db 481 |ACCTACCTCCTGAAATAGTGGTTCCTAAATCTGTTACCTTGGGAACGGTGTGGAAGT 540
Qy 541 |TCAAAGTTCAGAACTAAAGAACTGCTGCTGCTCTCTTACCTCTACAAAGAGAACT 600
Db 541 |TCAAAGTTCAGAACTAAAGAACTGCTGCTGCTCTCTTACCTCTACAAAGAGAACT 600
Qy 601 |GCTGCCATTTGCCGTGTAGCCAGCTCTCTCTGGATTTTACACTGCTGTGTAGATGAT 660
Db 601 |GCTGCCATTTGCCGTGTAGCCAGCTCTCTCTGGATTTTACACTGCTGTGTAGATGAT 660
Qy 661 |GAGCTCTGTGTGAGGCCATTAGCCAAACAAACCCAGGAGCCAGTTTATGATGTTGA 720
Db 661 |GAGCTCTGTGTGAGGCCATTAGCCAAACAAACCCAGGAGCCAGTTTATGATGTTGA 720
Qy 721 |GACAAGACCAAGATCTGGCAATTCCTTGTGCTCTAATATGAGAAATAGTTCTCCAATTA 780
Db 721 |GACAAGACCAAGATCTGGCAATTCCTTGTGCTCTAATATGAGAAATAGTTCTCCAATTA 780
Qy 781 |GCCAAGATGAACTCTATGAGACATCACCAGATCTTCTCCCTCTGAGCCGACAGGAG 840
Db 781 |GCCAAGATGAACTCTATGAGACATCACCAGATCTTCTCCCTCTGAGCCGACAGGAG 840
Qy 841 |GAGGAGCACTGACACAGAGGAGAGCAGGCTCTCAATCAAGCAGTGTATGACACGAC 900
Db 841 |GAGGAGCACTGACACAGAGGAGAGCAGGCTCTCAATCAAGCAGTGTATGACACGAC 900
Qy 901 |TCCATATCTTTGACACAGCTTTTGGCCAGGAGCTTACAAACGTTTTCATCAACAGCAGG 960
Db 901 |TCCATATCTTTGACACAGCTTTTGGCCAGGAGCTTACAAACGTTTTCATCAACAGCAGG 960
Qy 961 |AGTGGCTGGGGTGTCTCTGGACACCTTTGGCGTGGCTGCTTTTATGGCCACTTTGAGC 1020
Db 961 |AGTGGCTGGGGTGTCTCTGGACACCTTTGGCGTGGCTGCTTTTATGGCCACTTTGAGC 1020
Qy 1021 |TGTTCGAAGTCTCTTAGCCCATGTGCTGATGTTGACAGCTTGTGATGTCAAGGCACAG 1080
Db 1021 |TGTTCGAAGTCTCTTAGCCCATGTGCTGATGTTGACAGCTTGTGATGTCAAGGCACAG 1080
Qy 1081 |ACGCCACTTTTCACTGCTGATGCGCATCTGGACTGTGATGCTGTGCTTTTGGAA 1140
Db 1081 |ACGCCACTTTTCACTGCTGATGCGCATCTGGACTGTGATGCTGTGCTTTTGGAA 1140
Qy 1141 |GCTGGTGCCTCTCTCGTGTAGCATCTACAACTGTTCTCCCGTGTCTACAGCTGCC 1200
Db 1141 |GCTGGTGCCTCTCTCGTGTAGCATCTACAACTGTTCTCCCGTGTCTACAGCTGCC 1200
Qy 1201 |CGTGATGGTCTGCTTCTATCTGAGGAGCTCTAGACCATGCTGTCAGAGGCCACGTC 1260
Db 1201 |CGTGATGGTCTGCTTCTATCTGAGGAGCTCTAGACCATGCTGTCAGAGGCCACGTC 1260
Qy 1261 |AAAGCTAAACTACAGTCTGGGATCAAAATAGCTTTCATGTTCTGGCCCCCTCTATTTG 1320
```

```
Db 1261 |AAAGCTAAACTACAGTCTGGGATCAAAATAGCTTTCATGTTCTGGCCCCCTCTATTTG 1320
Qy 1321 |GCCGAGTCTACGGGACCTGACCTGTTTCCGCTGCTTTTCTCCACGGGCGAGACCT 1380
Db 1321 |GCCGAGTCTACGGGACCTGACCTGTTTCCGCTGCTTTTCTCCACGGGCGAGACCT 1380
Qy 1381 |GACTACAACTGACCTGACCCAGGCGCTATTTGGCTCGTGTCCCAAGACCCCGCACCTT 1440
Db 1381 |GACTACAACTGACCTGACCCAGGCGCTATTTGGCTCGTGTCCCAAGACCCCGCACCTT 1440
Qy 1441 |GAAATCTGCTCCATCAATAATTTGAGCCAGATATATCCAGCTGTTAATCGATTTGGT 1500
Db 1441 |GAAATCTGCTCCATCAATAATTTGAGCCAGATATATCCAGCTGTTAATCGATTTGGT 1500
Qy 1501 |GCTAATATCTACCTTCCATCTCTCTCCCTTGACCTGACCTCAAGATGATAAAGCAT 1560
Db 1501 |GCTAATATCTACCTTCCATCTCTCTCCCTTGACCTGACCTCAAGATGATAAAGCAT 1560
Qy 1561 |GCATTGCTCTACAGCCGAGCCACTCTCCAGCTCACTTCTATCAAGGTCCGTTTAGTC 1620
Db 1561 |GCATTGCTCTACAGCCGAGCCACTCTCCAGCTCACTTCTATCAAGGTCCGTTTAGTC 1620
Qy 1621 |GTCCGAGAGCTTGTGCGAGCTGCGCAGCCACCAAGCCATCAACAGCTGGATATTCCT 1680
Db 1621 |GTCCGAGAGCTTGTGCGAGCTGCGCAGCCACCAAGCCATCAACAGCTGGATATTCCT 1680
Qy 1681 |CCCATGTTGATTTAGCTTACCTAAACACCACTGTAATCTTGCACTCTCCAGGACTTA 1740
Db 1681 |CCCATGTTGATTTAGCTTACCTAAACACCACTGTAATCTTGCACTCTCCAGGACTTA 1740
Qy 1741 |TGATGCTCCGAAACACCTGCGGACTCACCTAGCTGGAGAGCATTCAGGCTCATCCA 1800
Db 1741 |TGATGCTCCGAAACACCTGCGGACTCACCTAGCTGGAGAGCATTCAGGCTCATCCA 1800
Qy 1801 |CTTACCTGAGCTGCTCTCTCTATTTATCTCCCAATAAAATTTCTCCAGAAAATAAGTA 1860
Db 1801 |CTTACCTGAGCTGCTCTCTCTATTTATCTCCCAATAAAATTTCTCCAGAAAATAAGTA 1860
Qy 1861 |AAAAAAAAAAAAAAAAAAAA 1878
Db 1861 |AAAAAAAAAAAAAAAAAAAA 1878
```

RESULT 2

```
US-10-207-175-13
; Sequence 13, Application US/10207175
; Publication No. US20030157508A1
; GENERAL INFORMATION:
; APPLICANT: Ebner et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ74P1
; CURRENT APPLICATION NUMBER: US/10/207,175
; PRIOR FILING DATE: 2002-07-30
; PRIOR APPLICATION NUMBER: PCT/US01/01395
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: US 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/186,350
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 1878
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-207-175-13
```

```
Query Match 100.0%; Score 1878; DB 12; Length 1878;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 CATGATTACCCAGCTTGCCACGAGGTAGAAAACGTGAATTTGGTGGATCGTTATGTG 60
DB |||||
1 CATGATTACCCAGCTTGCCACGAGGTAGAAAACGTGAATTTGGTGGATCGTTATGTG 60
QY 61 AGTAAGAACCCAGCTTAATGGGATCTTTATCTTACTGCAACCCACCTGATCTATGTGGAG 120
DB |||||
61 AGTAAGAACCCAGCTTAATGGGATCTTTATCTTACTGCAACCCACCTGATCTATGTGGAG 120
QY 121 GCTTCAGGTGACGCCGGAAGAAACATGGATTGGCACTCCATCACAATTCGCACTGTGGAG 180
DB |||||
121 GCTTCAGGTGACGCCGGAAGAAACATGGATTGGCACTCCATCACAATTCGCACTGTGGAG 180
QY 181 AAGTTACCCATCACTAGCCTGGGTGTTCCCTGACCCCTCGCTGCAAGAAATTTCCGGGTG 240
DB |||||
181 AAGTTACCCATCACTAGCCTGGGTGTTCCCTGACCCCTCGCTGCAAGAAATTTCCGGGTG 240
QY 241 GCCCAGCTTTGTTTAGATTCTGACCTTGTGTGCCATGAGTTTATATTTCACTGCTCAAG 300
DB |||||
241 GCCCAGCTTTGTTTAGATTCTGACCTTGTGTGCCATGAGTTTATATTTCACTGCTCAAG 300
QY 301 CTTTCTCAGCCAGCATTTACCTGAAGATCTTTATGCTTTTCTTTATATATCCCAATCCTCA 360
DB |||||
301 CTTTCTCAGCCAGCATTTACCTGAAGATCTTTATGCTTTTCTTTATATCCCAATCCTCA 360
QY 361 AAAGAGATGAGGAAAGTGGATGGAACCTGATTGACCCCAATATCAGACTTTGGGCGTATG 420
DB |||||
361 AAAGAGATGAGGAAAGTGGATGGAACCTGATTGACCCCAATATCAGACTTTGGGCGTATG 420
QY 421 GGNATACCCACAGAACTGGACCATTAACAGATGCCAACAGAACTATGAGATATCCAGC 480
DB |||||
421 GGAATACCCACAGAACTGGACCATTAACAGATGCCAACAGAACTATGAGATATCCAGC 480
QY 481 ACCTACCTCCTCAATAGTGGTTCCTAAATCTGTTACCTTGGGAACGGTGTGAAGT 540
DB |||||
481 ACCTACCTCCTCAATAGTGGTTCCTAAATCTGTTACCTTGGGAACGGTGTGAAGT 540
QY 541 TCAAAGTTCAGAAAGTAAAGAACCTGTCCTGCTCTCTTACCTCTACAAAGAGAAACAAT 600
DB |||||
541 TCAAAGTTCAGAAAGTAAAGAACCTGTCCTGCTCTCTTACCTCTACAAAGAGAAACAAT 600
QY 601 GCTGCCATTTGCGCTGTAGCAGCCTCTCTCTGGAATTTTACACTGCTGTGTAGATGAT 660
DB |||||
601 GCTGCCATTTGCGCTGTAGCAGCCTCTCTCTGGAATTTTACACTGCTGTGTAGATGAT 660
QY 661 GAGCTCTGTGTGAGGCCATTAGCCAAACAAACCCAGGGAGCCAGTTTATGTTGTTA 720
DB |||||
661 GAGCTCTGTGTGAGGCCATTAGCCAAACAAACCCAGGGAGCCAGTTTATGTTGTTA 720
QY 721 GACACAGACCAAGATCTGGCAATTTCTTGTGCTCATATGAGAAATAGTTCTCCAATTA 780
DB |||||
721 GACACAGACCAAGATCTGGCAATTTCTTGTGCTCATATGAGAAATAGTTCTCCAATTA 780
QY 781 GCCAAGATGAACCTCATGAGCATCAACAGATCTTCTCCCTCTGAGCCCGCAAGAGGAG 840
DB |||||
781 GCCAAGATGAACCTCATGAGCATCAACAGATCTTCTCCCTCTGAGCCCGCAAGAGGAG 840
QY 841 GAGGAGACACTGACACAGAGGAGAGCAGGCTCTCAATCAAGCAGTGTATGACACGAC 900
DB |||||
841 GAGGAGACACTGACACAGAGGAGAGCAGGCTCTCAATCAAGCAGTGTATGACACGAC 900
QY 901 TCCTATATTTGGACCACTTTTGGCCAGGAGCGTTACAAACGTTTTCATCAACACAGG 960
DB |||||
901 TCCTATATTTGGACCACTTTTGGCCAGGAGCGTTACAAACGTTTTCATCAACACAGG 960
QY 961 AFTGGCTGGGTTTCTGCGGACACCCCTTGGCTGCTTCTTATGGCCACTTGGAGC 1020
DB |||||
961 AFTGGCTGGGTTTCTGCGGACACCCCTTGGCTGCTTCTTATGGCCACTTGGAGC 1020
QY 1021 TGTTCGAACTCTCTAGCCCATGCTGCTGATGTTGACAGCTTGTATGATCAAGGACAG 1080
DB |||||
1021 TGTTCGAACTCTCTAGCCCATGCTGCTGATGTTGACAGCTTGTATGATCAAGGACAG 1080
QY 1081 ACGCCACTTTTCACTGCTGCTAGTCATGGCCATCTGGAATGCTACGTGTGCTTTTGGAA 1140

DB |||||
1081 ACGCCACTTTTCACTGCTGCTAGTCATGGCCATCTGGACTGTGTACTGTCTTTGGAA 1140
QY 1141 GCTGGTCCCTCTCTGGTGGTAGCATCTACAACTGTTCTCCGTGCTCACAGTGGC 1200
DB |||||
1141 GCTGGTCCCTCTCTGGTGGTAGCATCTACAACTGTTCTCCGTGCTCACAGTGGC 1200
QY 1201 CGTGTAGTGGTGTGTTGCTATCTCGAGGAGCTCTTAGACCATGGTGCAGAGGCCAACGTC 1260
DB |||||
1201 CGTGTAGTGGTGTGTTGCTATCTCGAGGAGCTCTTAGACCATGGTGCAGAGGCCAACGTC 1260
QY 1261 AAAGCTAAACTACAGTCTGGGCATCAACATAGCTTCAATGTTCTGGCCCCCTCTATTG 1320
DB |||||
1261 AAAGCTAAACTACAGTCTGGGCATCAACATAGCTTCAATGTTCTGGCCCCCTCTATTG 1320
QY 1321 GCCGAGTCTACGGGACCTGGACTGTTTCCGCCCTGCTTTGCTCCACGGGCGAGACCT 1380
DB |||||
1321 GCCGAGTCTACGGGACCTGGACTGTTTCCGCCCTGCTTTGCTCCACGGGCGAGACCT 1380
QY 1381 GACTACACTGCACTGACCCAGGCGCTATTGGCTCGTGTCCCAAGACCCCGCACCTCTT 1440
DB |||||
1381 GACTACACTGCACTGACCCAGGCGCTATTGGCTCGTGTCCCAAGACCCCGCACCTCTT 1440
QY 1441 GAAATCTGCTCCATCATATAATTTGTGAGCCAGAGTATATCCAGCTGTTAATCGATTTG 1500
DB |||||
1441 GAAATCTGCTCCATCATATAATTTGTGAGCCAGAGTATATCCAGCTGTTAATCGATTTG 1500
QY 1501 GCTAATATCTACCTTCCATCTCTCTCCCTTGACCTGACCTCAAGATGATAAGGCAT 1560
DB |||||
1501 GCTAATATCTACCTTCCATCTCTCTCCCTTGACCTGACCTCAAGATGATAAGGCAT 1560
QY 1561 GCATTGCTGTACAGCCGAGCGCTATCCAGCTGCTCTCTCCCTTGACCTGACCTCAAGATGATAAGGCAT 1620
DB |||||
1561 GCATTGCTGTACAGCCGAGCGCTATCCAGCTGCTCTCTCTCCCTTGACCTGACCTCAAGATGATAAGGCAT 1620
QY 1621 GTCCGAGAGCCTTGTGCGAGCTGCGCCAGCCACCAAGCCATCAACAGCTGGATTTCT 1680
DB |||||
1621 GTCCGAGAGCCTTGTGCGAGCTGCGCCAGCCACCAAGCCATCAACAGCTGGATTTCT 1680
QY 1681 CCATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
DB |||||
1681 CCATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
QY 1741 TCATGCTCCGAAACACCTGCGGAGCTCAGCTGAGAGCATTAAGAGTATTAAGTA 1800
DB |||||
1741 TCATGCTCCGAAACACCTGCGGAGCTCAGCTGAGAGCATTAAGAGTATTAAGTA 1800
QY 1801 CTTTACCTGAGCTGCT 1860
DB |||||
1801 CTTTACCTGAGCTGCT 1860
QY 1861 AAAAAAIAAAAAAAAAA 1878
DB |||||
1861 AAAAAAIAAAAAAAAAA 1878

RESULT 3

US-09-908-193-13
; Sequence 13, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHINKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
; APPLICANT: PADICARU, MURALIDHARA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 21402-062
; CURRENT APPLICATION NUMBER: US/09/908,193
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: 60/220,273
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/221,650

PRIOR FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: 60/221,233
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: 60/220,912
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/218,875
PRIOR FILING DATE: 2000-07-18
PRIOR APPLICATION NUMBER: 60/218,870
PRIOR FILING DATE: 2000-07-18
PRIOR APPLICATION NUMBER: 60/218,901
PRIOR FILING DATE: 2000-07-18
NUMBER OF SEQ ID NOS: 74
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 13
LENGTH: 1222
TYPE: DNA
ORGANISM: Unknown Organism
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: NOVX nucleic acid
with homology to ankryrin repeat-containing-
OTHER INFORMATION: (ASB-1) like protein
US-09-908-193-13

Query Match 59.5%; Score 1117.8; DB 10; Length 1222;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1119; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

730 CCAAGATCTGGCATTCCTTGTGCTCAATGAGAAATGCTTCAATTTAGCCAGATG 789
102 CCACAGATCTGGCATTCCTTGTGCTCAATGAGAAATGCTTCAATTTAGCCAGATG 161
790 AACCTCATGACATCCAGAGATCTTCTCCCTCCGAGCCGACAGGAGGAGGAC 849
162 AACCTCATGACATCCAGAGATCTTCTCCCTCCGAGCCGACAGGAGGAGGAC 221
850 ACTGACACAGAGAGAGAGGCTCTCAATCAAGCAGTGTATGACAAACGACTCTTACT 909
222 ACTGACACAGAGAGAGAGGCTCTCAATCAAGCAGTGTATGACAAACGACTCTTACT 281
910 TTGGACAGCTTTTGGCCAGAGAGCGTTACAAACGTTTCATCAACAGCAGGAGTGGCTGG 969
282 TTGGACAGCTTTTGGCCAGAGAGCGTTACAAACGTTTCATCAACAGCAGGAGTGGCTGG 341
970 GGTGTTCTGGGACACCTTGGCTGGCTTCTTATGGCCACTGAGCTGTTGCAA 1029
342 GGTGTTCTGGGACACCTTGGCTGGCTTCTTATGGCCACTGAGCTGTTGCAA 401
1030 GTCTCTTAGCCCATGCTGATGTTGACAGCTTGGATGTCAGGACAGAGCCACTT 1089
402 GTCTCTTAGCCCATGCTGATGTTGACAGCTTGGATGTCAGGACAGAGCCACTT 461
1090 TTCACTGCTGTAGTCATGATGCCATCTGGACTGTGATGCTGTTTGAAGCTGTGCC 1149
462 TTCACTGCTGTAGTCATGATGCCATCTGGACTGTGATGCTGTTTGAAGCTGTGCC 521
1150 TCTCCTGGTGTAGTCATGATGCCATCTGGACTGTGATGCTGTTTGAAGCTGTGCC 1209
522 TCTCCTGGTGTAGTCATGATGCCATCTGGACTGTGATGCTGTTTGAAGCTGTGCC 581
1210 GCTGTTGCTATCTCTGAGGAGCTCTTAGACCATGGTGCAGAGCCCAACGTCAGGCTAA 1269
582 GCTGTTGCTATCTCTGAGGAGCTCTTAGACCATGGTGCAGAGCCCAACGTCAGGCTAA 641
1270 CTACAGCTGTGGCATCAAAATAGTGTTCATGTTCTGGCCCTCTTATTTGGCCGAGTC 1329
642 CTACAGCTGTGGCATCAAAATAGTGTTCATGTTCTGGCCCTCTTATTTGGCCGAGTC 701
1330 TAGGGCACTGACCTGTTTCCGCTGTTTCTCCAGGGGACAGCCCTGACTACAC 1389
702 TAGGGCACTGACCTGTTTCCGCTGTTTCTCCAGGGGACAGCCCTGACTACAC 761
1390 TGCACTGACAGGAGCTTATGGCTGCTGTCAGGACCCGACCCCTCTTGAATCTGC 1449

762 TGCACTGACAGGAGCTTATGGCTGCTGTCCTCAAGACCCCGACCCCTCTTGAATCTGC 821
1450 CTCCATCATTAATGTGAGCCAGAGTATATCCAGCTGTTAATCGATTTTGGTGTATATC 1509
822 CTCCATCATTAATGTGAGCCAGAGTATATCCAGCTGTTAATCGATTTTGGTGTATATC 881
1510 TACCTTCCATCTCTCTCCCTTGACCTGACCTCAAGATGATAAAGGCATTGCAATGCTG 1569
882 TACCTTCCATCTCTCTCCCTTGACCTGACCTCAAGATGATAAAGGCATTGCAATGCTG 941
1570 CTACAGGCGGAGCCACTCCACGGTCACTTCTATCAAGTCCGTTTGTGTCGACAGA 1629
942 CTACAGGCGGAGCCACTCCACGGTCACTTCTATCAAGTCCGTTTGTGTCGACAGA 1001
1630 GCTTGTGTCAGGCTGGCCAGCCACAGCCATCAAGCCATCAAGCCATGATTTCTCCCATGTTG 1689
1002 GCTTGTGTCAGGCTGGCCAGCCACAGCCATCAAGCCATCAAGCCATGATTTCTCCCATGTTG 1061
1690 ATTAGCTACTTAAACACCAACTGTAATCTTCAGTCTCCCGAGGAACCTTATGATCCCTC 1749
1062 ATTAGCTACTTAAACACCAACTGTAATCTTCAGTCTCCCGAGGAACCTTATGATCCCTC 1121
1750 CGAAACACCTTGGGACTCAGTGTGAGAGGATTAAGCCCTCATCCACTTACCTGG 1809
1122 CGAAACACCTTGGGACTCAGTGTGAGAGGATTAAGCCCTCATCCACTTACCTGG 1181
1810 AGCTGCTCTCTGTTATTTATCTCCCAATATAAATTTCTCCAG 1850
1182 AGCTGCTCTCTGTTATTTATCTCCCAATATAAATTTCTCCAG 1222

RESULT 4

US-09-908-193-11
; Sequence 11, Application US/09908193
; Publication No. US2002012748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHIMKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
; APPLICANT: PADIGARU, MURALIDHARA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 21402-062
; CURRENT APPLICATION NUMBER: US/09/908,193
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: 60/220,273
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/221,650
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/221,233
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/220,912
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/218,875
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,870
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,901
; PRIOR FILING DATE: 2000-07-18
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 1069
; TYPE: DNA
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: NOVX nucleic acid
with homology to ankryrin repeat-containing-
; OTHER INFORMATION: (ASB-1) like protein
US-09-908-193-11

Query Match 49.6%; Score 931; DB 10; Length 1069;
Best Local Similarity 97.6%; Pred. No. 6e-291;

Matches 965; Conservative 0; Mismatches 0; Indels 24; Gaps 1;

QY 760 ATGAGATAGTCTCCAAATAGCAAGATGAACCTCATGACATACCAAGATCTTCTCC 819
DB |||||
QY 81 ATGAGATAGTCTCCAAATAGCAAGATGAACCTCATGACATACCAAGATCTTCTCC 140
DB |||||
QY 820 CTCCTCAGCCGACAAAGGAGGAGGAGCACTGACACAGAGAGAGAGGCTCTCAAT 879
DB |||||
QY 141 CTCCTCAGCCGACAAAGGAGGAGGAGCACTGACACAGAGAGAGAGGCTCTCAAT 200
DB |||||
QY 880 CAAGCAGTGTATGACAAAGCACTCTATCTTGGACAGCTTTGGCCGAGAGCGTTAC 939
DB |||||
QY 201 CAAGCAGTGTATGACAAAGCACTCTATCTTGGACAGCTTTGGCCGAGAGCGTTAC 260
DB |||||
QY 940 AAAGCTTTTCATCAACAGAGGAGTGTGGGGTTCCTGGGACACCTTTGGCTGGCT 999
DB |||||
QY 261 AAAGCTTTTCATCAACAGAGGAGTGTGGGGTTCCTGGGACACCTTTGGCTGGCT 320
DB |||||
QY 1000 GCTTCTTATGGCCACTTGAGCTGTTTGCAGTCTCTTGGCCCATGGTGTGATGTGAC 1059
DB |||||
QY 321 GCTTCTTATGGCCACTTGAGCTGTTTGCAGTCTCTTGGCCCATGGTGTGATGTGAC 380
DB |||||
QY 1060 AGCTTGGATGTCAAGGACAGAGCCACTTTTCACTGCTGTGATGATGGCCATCTGGAC 1119
DB |||||
QY 381 AGCTTGGATGTCAAGGACAGAGCCACTTTTCACTGCTGTGATGATGGCCATCTGGAC 440
DB |||||
QY 1120 TGTGTACGTGTGCTTTTGGAGCTGTGCTCTCTGCTGGTGGTGTGATCTACAACTGT 1179
DB |||||
QY 441 TGTGTACGTGTGCTTTTGGAGCTGTGCTCTCTGCTGGTGGTGTGATCTACAACTGT 500
DB |||||
QY 1180 TCTCCCGTGTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1239
DB |||||
QY 501 TCTCCCGTGTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 560
DB |||||
QY 1240 CATGTGTGAGAGCCAAAGCTCAAGCTAACTACAGTCTGGGCTCAAACTAGCTTCA 1299
DB |||||
QY 561 CATGTGTGAGAGCCAAAGCTCAAGCTAACTACAGTCTGGGCTCAAACTAGCTTCA 620
DB |||||
QY 1300 TGTTCCTGGCCCTCTATTTGGCCGAGTCTAGGGCAGCTGAGCTGTTTCGGCTGCTT 1359
DB |||||
QY 621 TGTTCCTGGCCCTCTATTTGGCCGAGTCTAGGGCAGCTGAGCTGTTTCGGCTGCTT 680
DB |||||
QY 1360 TTGCTCCAGGGGAGAGCCCTGACTACAACTGACCTGACGAGGCTTATGGCTCGTGT 1419
DB |||||
QY 681 TTGCTCCAGGGGAGAGCCCTGACTACAACTGACCTGACGAGGCTTATGGCTCGTGT 740
DB |||||
QY 1420 CCAAGACCCCGCACCTCTCTTGAATCTGCTCCATCAATTTGTGAGCCAGATATATC 1479
DB |||||
QY 741 CCAAGACCCCGCACCTCTCTTGAATCTGCTCCATCAATTTGTGAGCCAGATATATC 800
DB |||||
QY 1480 CAGCTGTTAATCGATTTGGTGTATATCTACCTTCCATCTCTCTCCCTTGACCTGACC 1539
DB |||||
QY 801 CAGCTGTTAATCGATTTGGTGTATATCTACCTTCCATCTCTCTCCCTTGACCTGACC 860
DB |||||
QY 1540 TCACAGATGATAAAGGCAATGCTGCTGCTACAGGCCG----- 1580
DB |||||
QY 861 TCACAGATGATAAAGGCAATGCTGCTGCTACAGGCCGAGGTGAGCTGTTCTCTT 920
DB |||||
QY 1581 -----AGCCATCCACGGTCACTTCTATCAAGGTCCGTTTGTGTCGAGAGCCTTG 1635
DB |||||
QY 921 GCTGTAGCCACTCCAGGTCACCTTCTATCACAGGTCCGTTTGTGCTCGCAGAGCCTTG 980
DB |||||
QY 1636 TGCCAGGCTGGCCAGGACACAGGCAATCAACAGCTGATATCTCCCATGTTGATAGC 1695
DB |||||
QY 981 TGCCAGGCTGGCCAGGACACAGGCAATCAACAGCTGATATCTCCCATGTTGATAGC 1040
DB |||||
QY 1696 TACCTAAAACACCACTGTAATCTTGCA 1724
DB |||||
QY 1041 TACCTAAAACACCACTGTAATCTTGCA 1069
DB |||||

RESULT 5
US-09-833-381-62

; Sequence 62, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq For Windows Version 3.0
; SEQ ID NO 62
; LENGTH: 522
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(522)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-62

Query Match 16.9%; Score 317; DB 10; Length 522;
Best Local Similarity 88.6%; Pred.No. 1.1e-91;
Matches 437; Conservative 0; Mismatches 37; Indels 19; Gaps 8;

QY 971 GGTTCCTGGGACACCCCTTGGCTGGCTGCTTATGGCCACTTGGCTGTTGCAAG 1030
DB |||||
QY 1031 TCCTCTTAGCCCATGGTGTGATGTTGACAGCTTGGATGATCAAGGCACAGAGCCACTTT 1090
DB |||||
QY 71 TCCTCTTAGCCCATGGTGTGATGTTGACAGCTTGGATGATCAAGGCACAGAGCCACTTT 130
DB |||||
QY 1091 TCCTCTTAGCCCATGGTGTGATGTTGACAGCTTGGATGATCAAGGCACAGAGCCACTTT 1150
DB |||||
QY 131 TCCTCTTAGCCCATGGTGTGATGTTGACAGCTTGGATGATCAAGGCACAGAGCCACTTT 190
DB |||||
QY 1151 TCCTCTTAGCCCATGGTGTGATGTTGACAGCTTGGATGATCAAGGCACAGAGCCACTTT 1210
DB |||||
QY 191 TCCTCTTAGCCCATGGTGTGATGTTGACAGCTTGGATGATCAAGGCACAGAGCCACTTT 250
DB |||||
QY 1211 TCCTCTTAGCCCATGGTGTGATGTTGACAGCTTGGATGATCAAGGCACAGAGCCACTTT 1267
DB |||||
QY 251 TCCTCTTAGCCCATGGTGTGATGTTGACAGCTTGGATGATCAAGGCACAGAGCCACTTT 310
DB |||||
QY 1268 AACTACC-AGTCTGGGCAATCAAACTAGCTTCAATGTTCTGGCCCTCTCTA-TTT 1319
DB |||||
QY 311 AACTACCAGTCTGGGCAATCAAACTAGCTTCAATGTTCTGGCCCTCTCTAATTTG 370
DB |||||
QY 1320 GGGCCAGTCTAGGGGCACTGGACTGTTTCGGCTGCTTTCCTCCAGGGG---CAGA 1376
DB |||||
QY 371 GGGCCAGTCTAGGGGCACTGGACTGTTTCGGCTGCTTTCCTCCAGGGGCAAGAA 430
DB |||||
QY 1377 CCCTGACTAC-AACTGCACTGACAGGGGCT---TATTGGCTGCTGCTCCAGAGACCCCGC 1431
DB |||||
QY 431 CCCTGACTACAAATGCACTGACAGGGGCTAATTTGGTTCGTTGTCNCAAGAACCCGN 490
DB |||||

RESULT 6
US-09-908-805B-24
; Sequence 24, Application US/09908805B
; Patent No. US20020147307A1
; GENERAL INFORMATION:
; APPLICANT: HILTON, Douglas J
; APPLICANT: ALEXANDER, Warren S
; APPLICANT: VINEY, Elizabeth M
; APPLICANT: WILLSON, Tracey A
; APPLICANT: RICHARDSON, Rachael T

GENERAL INFORMATION:
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Stolk, John A.
APPLICANT: Meagher, Madeleine J.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
FILE REFERENCE: 210121.471C12
CURRENT APPLICATION NUMBER: US/09/833,263
CURRENT FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 1093
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 28
LENGTH: 401
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(401)
OTHER INFORMATION: n = A,T,C or G
US-09-833-263-28

Query Match 3.7%; Score 69.4; DB 10; Length 401;
Best Local Similarity 52.4%; Pred. No. 2.5e-11;
Matches 176; Conservative 0; Mismatches 157; Indels 3; Gaps 1;
QY 360 AAAGAGATGAGGAAAGTGGATGGAACTGATTGACCCCAATATCAGACTTTGGGCGTAT 419
DB 52 AGAAAAGTTTAACTGGATGGAGACAGTTTACAATCCAGTGGAGAAATACAGGAGGCA 111
QY 420 GGAATATACCAACAGAACTGGACCAATACAGATGCCAAGAACTATGAGATATGCAG 479
DB 112 GGGCTTGGCCCAATCACCATTGGAGATACTTTTATTATAGTCTATGAGCTCTGCCA 171
QY 480 CACCTACCTCTGAAATAGTGTCTTAAATCTGTACCTTGGGAACGGTGTGGAAG 539
DB 172 CACTTACCTCTCTTTTGGTGTTCGGTATCGTGCTCANATGATGACCTCCGGAGAGT 231
QY 540 TTCAAAGTTTCAAGATAAGAAAGTGTCCCTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 599
DB 232 TGCACTTTTATAGTCCGGAATCGAATTCAGTGTCTGTATGAGTTCATCCAGAAATAA 291
QY 600 TGCTGCATTTGCGCTGTAGCCAGCTCTCTCTGGATT---TTACACTGCTGTGTAGA 656
DB 292 GACGGTCATTGTCGGTTCAGTCCAGCTCTTGTTCGGTATGATGGGAAACGAATAAGA 351
QY 657 TGATGAGCTCTTTGGAGGCCATTAGCCAAACAAA 692
DB 352 TGATGAGAAATATCTCGATGTTATCAGGGAGACTAA 387

RESULT 12
US-10-025-380-28
Sequence 28, Application US/10025380
Publication No. US20020182191A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tongtong
APPLICANT: Jiang, Yuelu
APPLICANT: Smith, Carole L.
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Skeiky, Yasir A. W.
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS

TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
FILE REFERENCE: 210121.471C14
CURRENT APPLICATION NUMBER: US/10/025,380
CURRENT FILING DATE: 2001-12-19
NUMBER OF SEQ ID NOS: 1129
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 28
LENGTH: 401
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 212
OTHER INFORMATION: n = A,T,C or G
US-10-025-380-28

Query Match 3.7%; Score 69.4; DB 13; Length 401;
Best Local Similarity 52.4%; Pred. No. 2.5e-11;
Matches 176; Conservative 0; Mismatches 157; Indels 3; Gaps 1;
QY 360 AAAGAGATGAGGAAAGTGGATGGAACTGATTGACCCCAATATCAGACTTTGGGCGTAT 419
DB 52 AGAAAAGTTTAACTGGATGGAGACAGTTTACAATCCAGTGGAGAAATACAGGAGGCA 111
QY 420 GGAATATACCAACAGAACTGGACCAATACAGATGCCAAGAACTATGAGATATGCAG 479
DB 112 GGGCTTGGCCCAATCACCATTGGAGATACTTTTATTATAGTCTATGAGCTCTGCCA 171
QY 480 CACCTACCTCTGAAATAGTGTCTTAAATCTGTACCTTGGGAACGGTGTGGAAG 539
DB 172 CACTTACCTCTCTTTTGGTGTTCGGTATCGTGCTCANATGATGACCTCCGGAGAGT 231
QY 540 TTCAAAGTTTCAAGATAAGAAAGTGTCCCTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 599
DB 232 TGCACTTTTATAGTCCGGAATCGAATTCAGTGTCTGTATGAGTTCATCCAGAAATAA 291
QY 600 TGCTGCATTTGCGCTGTAGCCAGCTCTCTCTGGATT---TTACACTGCTGTGTAGA 656
DB 292 GACGGTCATTGTCGGTTCAGTCCAGCTCTTGTTCGGTATGATGGGAAACGAATAAGA 351
QY 657 TGATGAGCTCTTTGGTGGAGGCCATTAGCCAAACAAA 692
DB 352 TGATGAGAAATATCTCGATGTTATCAGGGAGACTAA 387

RESULT 13
US-09-919-039-133
Sequence 133, Application US/09919039
Publication No. US20030108871A1
GENERAL INFORMATION:
APPLICANT: Kaser, Matthew R.
TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
FILE REFERENCE: PA-0035 US
CURRENT APPLICATION NUMBER: US/09/919,039
CURRENT FILING DATE: 2002-09-09
PRIOR APPLICATION NUMBER: 60/222,113
PRIOR FILING DATE: 2000-07-28
NUMBER OF SEQ ID NOS: 401
SOFTWARE: PERL Program
SEQ ID NO 133
LENGTH: 3482
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. US20030108871A1 1823159CBI
US-09-919-039-133

Query Match 3.6%; Score 67.2; DB 11; Length 3482;
Best Local Similarity 52.1%; Pred. No. 5.7e-10;
Matches 175; Conservative 0; Mismatches 158; Indels 3; Gaps 1;
QY 360 AAAGAGATGAGGAAAGTGGATGGAACTGATTGACCCCAATATCAGACTTTGGGCGTAT 419

Search completed: December 1, 2003, 13:44:05
Job time : 620 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2003, 07:34:53 ; Search time 68 Seconds
(without alignments)
1048.062 Million cell updates/sec

Title: US-09-941-831A-20
Perfect score: 2343
Sequence: 1 MRESGWKLPISTDFGRMG1.....QAINQLDIPPLISYLKQL 449

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A. Geneseq 19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2343	100.0	449	22	Human serine/threose
2	2343	100.0	449	22	Human HIBGJ89 seri
3	2343	100.0	449	22	Human secreted pro
4	2331	99.5	580	22	Human secreted pro
5	1890	72.1	360	23	Human NOV5b protei
6	1635	69.8	318	23	Human polypeptide
7	1629	69.5	351	23	Human NOV5a protei
8	642.5	27.4	370	23	Novel human protei
9	499	21.3	337	24	Human cytoskeleton

10	499	21.3	341	22	AAU20362	Human secreted pro
11	499	21.3	343	22	AAU20580	Human secreted pro
12	480	20.5	350	19	AAW62621	Mus musculus SOCS7
13	423	18.1	629	22	AAW73229	Human phosphatase
14	392.5	16.8	482	23	ABP62857	Human polypeptide
15	360	15.4	639	22	ABB65465	Drosophila melanog
16	360	15.4	676	22	ABB59964	Drosophila melanog
17	360	15.4	676	22	ABB66818	Drosophila melanog
18	297	12.7	393	22	AAW93364	Human protein sequ
19	281	12.0	665	22	AAW78876	Human protein SQ
20	281	12.0	673	23	AAE35082	Human kinase and p
21	275.5	11.8	858	22	ABG09695	Novel human diagno
22	273	11.7	705	22	AAW79860	Human protein SQ
23	270.5	11.5	278	21	AAW53886	A suppressor of cy
24	264.5	11.3	278	22	AAW95322	Human protein sequ
25	264	11.3	621	18	AAW36451	Human tyrosine pho
26	261.5	11.2	276	22	AAW39002	Human polypeptide
27	261.5	11.2	329	23	AAW50585	Benign prostatic h
28	256.5	10.9	299	23	ABP41229	Human ovarian anti
29	246	10.5	323	23	ABB05439	Human C/SKARP-1 pr
30	244.5	10.4	690	22	ABW11804	Human ASB-3 protei
31	243.5	10.4	518	21	AAW53889	A suppressor of cy
32	243.5	10.4	518	22	AAW92538	Human protein sequ
33	241	10.3	342	22	ABW11919	Human cytokine sig
34	241	10.3	342	22	AAW79442	Human protein SQ
35	235	10.0	407	23	ABP69199	Human polypeptide
36	235	10.0	435	22	AAW66710	Human cell growth
37	235	10.0	435	22	AAW93879	Human protein sequ
38	235	10.0	435	23	ABP69654	Human polypeptide
39	235	10.0	627	23	AAO17136	Human cancer cell
40	233.5	10.0	1377	22	ABG08072	Novel human diagno
41	231	9.9	362	21	AAW51508	Human secreted pro
42	230	9.8	293	21	AAW51425	Human secreted pro
43	230	9.8	293	21	AAW51427	Human secreted pro
44	228	9.7	784	22	ABW62206	Drosophila melanog
45	224.5	9.6	447	22	ABW53030	Cell structure and

ALIGNMENTS

RESULT 1

AAW85477
ID AAW85477 standard; Protein; 449 AA.

XX AAW85477;

AC AAW85477;

XX 25-SEP-2001 (first entry)

XX Human serine/threonine phosphatase (clone ID HIBGJ89).

XX Serine/threonine phosphatase; neotropic; neuroprotective; cytostatic;
XX dermatological; immunosuppressive; antiinflammatory; antibacterial;
XX anti-HIV; antiparkinsonian; antitickling; antianemic; antiarthritic;
XX antirheumatic; virucide; hepatotropic; cerebrotropic; vulnerary;
XX antiinflammatory; nephrotropic; gene therapy; vaccine.

OS Homo sapiens.

XX WO200155388-A1.

XX 02-AUG-2001.

PD 17-JAN-2001; 2001WO-US01395.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 02-MAR-2000; 2000US-0186350.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;

DR WPI: 2001-476208/51.
DR N-PSDB; AAH46857.
XX Novel proteins of serine/threonine phosphatase family, useful for
PT diagnosing, treating, preventing and/or prognosing disorders related to
PT the proteins, including cancer, immune response and neuronal disorders
PT
XX
XX
PS Claim 11; Page 432-433; 441pp; English.
XX
CC The invention provides serine/threonine phosphatase family polypeptides
CC and polynucleotides encoding them. The polypeptides can be expressed by
CC standard recombinant methodology. The polypeptides, polynucleotides and
CC antibodies are useful for diagnosis, prognosis, prevention and treatment
CC of neurodegenerative disorders, immune system disorders, autoimmune
CC diseases, allergic reactions, infectious diseases, hyperproliferative
CC disorders, renal disorders, cardiovascular disorders, cerebrovascular
CC disorders, respiratory disorders, endocrine disorders, gastrointestinal
CC disorders and also muscular, reproductive disorders (see AAH46846 for a
CC detailed description of the diseases that can be treated). They are also
CC useful as immune system enhancers, immunosuppressive agents, stimulator
CC of B-cell responsiveness to pathogens, activator of T-cells, to induce
CC higher affinity antibodies, as an agent to increase serum immunoglobulin
CC concentrations, to accelerate recovery of immunocompromised individuals,
CC to boost immunoresponsiveness in aged populations and/or neonates, as
CC regulator of antigen presentation, as a means to induce tumour
CC proliferation in pathologies e.g. AIDS, and/or common variable immuno
CC deficiency. The polypeptides and polynucleotides are useful to prevent
CC skin aging, for preventing hair loss, to stimulate growth and
CC differentiation of hemopoietic cells and bone marrow cells, for
CC supporting cell culture of primary tissues, to modulate mammalian
CC characteristics such as body weight, height, eye color, hair color and
CC skin, to modulate mammalian metabolism to change a mammal's mental or
CC physical state, and as food additive or preservative. The present
CC sequence represents a human serine/threonine phosphatase.
XX
SQ Sequence 449 AA;
Query Match 100.0%; Score 2343; DB 22; Length 449;
Best Local Similarity 100.0%; Pred. No. 2.6e-252;
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRESGKGLDPTDFGRMGIPNRRNWTITDANRNYEICSTYPPPIVVPKSVTLGTVVGSSK 60
DB 1 MRESGKGLDPTDFGRMGIPNRRNWTITDANRNYEICSTYPPPIVVPKSVTLGTVVGSSK 60
QY 61 FRSKRPVPLSYLYKENNAICRCSQPLSGFYTRCVDDLELLEAISQTNPGSQFMVYVDT 120
DB 61 FRSKRPVPLSYLYKENNAICRCSQPLSGFYTRCVDDLELLEAISQTNPGSQFMVYVDT 120
QY 121 RPKIWHFLVLMRIVLQAKNLMIDTKIFSLLOPKDEEDTTEBKQALNQAVYDNDYS 180
DB 121 RPKIWHFLVLMRIVLQAKNLMIDTKIFSLLOPKDEEDTTEBKQALNQAVYDNDYS 180
QY 181 TLDQLLRQERYKRFINSRSGWVGPTPLRLASVGHLSCLQVLHAGADVSLDVKQAQTP 240
DB 181 TLDQLLRQERYKRFINSRSGWVGPTPLRLASVGHLSCLQVLHAGADVSLDVKQAQTP 240
QY 241 LFTAVSHGLDCVRVLEAGAPGSGIYNNCSPLVTAARDGAVAILQELLDHGAENAVKA 300
DB 241 LFTAVSHGLDCVRVLEAGAPGSGIYNNCSPLVTAARDGAVAILQELLDHGAENAVKA 300
QY 301 KLPVWASNTASCSGPVLAAYVGHLDPCFRLLHLHGADPDYNTDQGLLARVPRPTLLBI 360
DB 301 KLPVWASNTASCSGPVLAAYVGHLDPCFRLLHLHGADPDYNTDQGLLARVPRPTLLBI 360
QY 361 CLHHCNPEYIQLLIDFGANILPSSLTDLTSDQDKGIALLLQARATPRSLLSQVRLVVR 420
DB 361 CLHHCNPEYIQLLIDFGANILPSSLTDLTSDQDKGIALLLQARATPRSLLSQVRLVVR 420
QY 421 RALCQAGQQAQINQLDIPPMILSYLKHQL 449
DB 421 RALCQAGQQAQINQLDIPPMILSYLKHQL 449

RESULT 2
AAG77803
ID AAG77803 standard; Protein; 449 AA.
XX
AC AAG77803;
XX
DT 03-DEC-2001 (first entry)
XX
DE Human HIBCU89 serine/threonine phosphatase protein sequence.
XX
KW Human; HIBCU89; serine/threonine phosphatase; PSPase; vaccine;
KW gene therapy; PSPase expression; PSPase expression; PSPase modulation;
KW immune disorder; autoimmune disorder; Wiscott-Aldrich syndrome;
KW Chediak-Higashi syndrome; Hashimoto's thyroiditis; multiple sclerosis;
KW inflammation; Crohn's disease; inflammatory bowel disease; appendicitis;
KW rheumatoid arthritis; cellular proliferative disorder; lymphoma;
KW lung cancer; intestinal cancer; cardiovascular disorder; aneurysm;
KW Scimitar syndrome; Ebstein's anomaly.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 29..35 /note= "Immunogenic region 1"
FT Region 59..66 /note= "Immunogenic region 2"
FT Region 106..112 /note= "Immunogenic region 3"
FT Region 154..169 /note= "Immunogenic region 4"
FT Region 175..180 /note= "Immunogenic region 5"
FT Region 185..201 /note= "Immunogenic region 6"
FT Region 336..344 /note= "Immunogenic region 7"
FT Region 391..396 /note= "Immunogenic region 8"
XX
PN WO200164703-A1.
XX
PD 07-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US06256.
XX
PR 02-MAR-2000; 2000US-0186350.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ebner R, Ruben SM;
XX
DR WPI; 2001-530113/58.
DR N-PSDB; AAH78730.
XX
PT Nucleic acids encoding serine/threonine phosphatase polypeptides,
PT useful for preventing, diagnosing and/or treating, e.g. Crohn's
PT disease, lung cancer and Scimitar syndrome -
XX
PS Claim 11; Page 321-323; 335pp; English.
XX
CC The present sequence represents the human HIBCU89 protein which is
CC claimed in the invention. The invention comprises novel human
CC serine/threonine phosphatase (PSPase) polypeptides and polynucleotides.
CC The PSPase polynucleotides and polypeptides of the invention may be used
CC in the prevention (vaccine), diagnosis and treatment (gene therapy) of
CC diseases associated with inappropriate PSPase expression. The PSPase
CC polynucleotides of the invention may be used as DNA probes to detect and
CC quantitate the presence of similar nucleic acids in samples. The PSPase
CC polypeptides may be used as antigens in the production of antibodies
CC against the PSPase polypeptides and in assays to identify modulators of
CC PSPase expression and activity. The anti-PSPase antibodies and

antagonists may also be used to down regulate expression and activity, the anti-PSpase antibodies may also be used as diagnostic agents for detecting the presence of PSpase polypeptides in samples. Disorders that may be prevented, diagnosed and/or treated by the invention are: immune/autoimmune disorders (e.g. Wiscott-Aldrich syndrome, Chediak-Higashi syndrome, Hashimoto's thyroiditis and multiple sclerosis); inflammatory conditions (e.g. Crohn's disease, inflammatory bowel disease, appendicitis and rheumatoid arthritis); cellular proliferative disorders (e.g. lymphoma, lung and intestinal cancers); and cardiovascular disorders (e.g. Scimitar syndrome, Ebstein's anomaly and aneurysm).

XX SQ Sequence 449 AA;

Query Match 100.0%; Score 2343; DB 22; Length 449;

Best Local Similarity 100.0%; Pred. No. 2.6e-252;

Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MRESGKWLIDPISDFGRMGIPNPNWTITDANRNYEICSTYPPPIVVPKSVTLGTVVGSSK 60
1 MRESGKWLIDPISDFGRMGIPNPNWTITDANRNYEICSTYPPPIVVPKSVTLGTVVGSSK 60
61 FRSKERPVLVLYKNNAAICRCSQPLSGFYTRCVDDLELLEAISQTNPGSQFMYVVD 120
61 FRSKERPVLVLYKNNAAICRCSQPLSGFYTRCVDDLELLEAISQTNPGSQFMYVVD 120
121 RPKIWHFLVLMRIVLQAKNLMIDITKIFSLQDPKEEDTTEKQALNQAQVVDNSY 180
121 RPKIWHFLVLMRIVLQAKNLMIDITKIFSLQDPKEEDTTEKQALNQAQVVDNSY 180
181 TLDQLLRQERYKRFINSRSGWGPVPTPLRLAASYGHLSCLOVLLAHGADVDSLVDKAQTP 240
181 TLDQLLRQERYKRFINSRSGWGPVPTPLRLAASYGHLSCLOVLLAHGADVDSLVDKAQTP 240
241 LFTAVSHGLDCVRVLEAGAPGGSIYNNCSFVLTAARDGAVAILQELLDHGAENYKA 300
241 LFTAVSHGLDCVRVLEAGAPGGSIYNNCSFVLTAARDGAVAILQELLDHGAENYKA 300
301 KLPVWASNTASCGPLVLAAYVGHLDPCFRLLLLHAGADPDYNTDQGLLARVPRPTLLI 360
301 KLPVWASNTASCGPLVLAAYVGHLDPCFRLLLLHAGADPDYNTDQGLLARVPRPTLLI 360
361 CLHNCPEYIQLLIDFGANILPSLSLDTTSQDDKGIALLQARATPSLLSQRLVVR 420
361 CLHNCPEYIQLLIDFGANILPSLSLDTTSQDDKGIALLQARATPSLLSQRLVVR 420
421 RALCQAGQPOAINQLDIPMLISYLKHQL 449
421 RALCQAGQPOAINQLDIPMLISYLKHQL 449

RESULT 3

AAU20663

ID AAU20663 standard; Protein; 449 AA.

XX AC AAU20663;

XX AC AAU20663;

DT 04-DEC-2001 (first entry)

XX XX Human secreted protein, Seq ID No 655.

Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina; rheumatoid arthritis; antiarteriosclerotic; cardiac; vascular; cerebroprotective; thrombolytic; antimicrobial; ophthalmological; cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer; multiple sclerosis; cancer; hyperproliferative disorder; infection; Gaucher's disease; neurological disease; cerebrovascular disorder; thrombosis; wound healing.

XX OS Homo sapiens.

XX XX WO200155326-A2.

PN WO200155326-A2.

17-JAN-2001; 2001WO-US01347.
31-JAN-2000; 2000US-0179065.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-451931/48.
N-PSDB; AAS33372.
New nucleic acids and polypeptides, useful for diagnosing, preventing or treating medical conditions -
Claim 11; SEQ ID No 655; 753pp; English.
The invention relates to novel isolated nucleic acid molecules (I) encoding human secreted proteins (II). (I) and (II) are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, in goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate expression of secreted proteins. (I) and complementary sequences may also be used as DNA probes in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate the presence of similar nucleic acid sequences in samples, and so which patients may be in need of restorative therapy. (II) may also be used as antigens in the production of antibodies and in assays to identify modulators (agonists and antagonists) of the expression and activity of the secreted proteins. The anti-(II) antibodies and antagonists may also be used to down regulate expression and activity of (II). The anti-(II) antibodies may also be used as diagnostic agents for detecting the presence of (II) in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The disorders include for example: immune/autoimmune diseases (e.g. HIV (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis and multiple sclerosis), cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of the breast or liver, Sezary syndrome and Gaucher's disease), neurological diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/cerebrovascular disorders (e.g. cardiac arrest, tachycardia, angina and thrombosis), infections caused by bacteria, viruses and fungi and ocular disorders (e.g. corneal infections). (I) and (II), agonists, antagonists and antibodies can also be used to promote wound healing, maintain organs before transplantation, and support cell culture of primary tissues. AAU20342-AAU20666 represent human secreted protein amino acid sequences, and related sequences of the invention.
Note: The sequence data for this patent did not appear in the printed specification but was obtained in electronic format directly from WIPO at: ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 449 AA;

Query Match 100.0%; Score 2343; DB 22; Length 449;

Best Local Similarity 100.0%; Pred. No. 2.6e-252;

Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRESGKWLIDPISDFGRMGIPNPNWTITDANRNYEICSTYPPPIVVPKSVTLGTVVGSSK 60

Db 1 MRESGKWLIDPISDFGRMGIPNPNWTITDANRNYEICSTYPPPIVVPKSVTLGTVVGSSK 60

QY 61 FRSKERPVLVLYKNNAAICRCSQPLSGFYTRCVDDLELLEAISQTNPGSQFMYVVD 120

Db 61 FRSKERPVLVLYKNNAAICRCSQPLSGFYTRCVDDLELLEAISQTNPGSQFMYVVD 120

QY 121 RPKIWHFLVLMRIVLQAKNLMIDITKIFSLQDPKEEDTTEKQALNQAQVVDNSY 180

Db 121 RPKIWHFLVLMRIVLQAKNLMIDITKIFSLQDPKEEDTTEKQALNQAQVVDNSY 180

QY 181 TLDQLLRQERYKRFINSRSGWGPVPTPLRLAASYGHLSCLOVLLAHGADVDSLVDKAQTP 240

Db 181 TLDQLLRQERYKRFINSRSGWGPVPTPLRLAASYGHLSCLOVLLAHGADVDSLVDKAQTP 240

QY 241 LFTAVSHGHLDVRLVLEAGSPGSIYNNCSFVLTAARDGAVAILQELLDHGAENVKA 300
 DB 241 LFTAVSHGHLDVRLVLEAGSPGSIYNNCSFVLTAARDGAVAILQELLDHGAENVKA 300
 QY 301 KLPVWASNIASCSGPLYLAAYVGHLDVRLVLEAGSPGSIYNNCSFVLTAARDGAVAILQELLDHGAENVKA 360
 DB 301 KLPVWASNIASCSGPLYLAAYVGHLDVRLVLEAGSPGSIYNNCSFVLTAARDGAVAILQELLDHGAENVKA 360
 QY 361 CLHNCPEPIYQLLDVFGANIYLPVLSLDLTSQDDKGIALLQARATPRSLLSQVRLVVR 420
 DB 361 CLHNCPEPIYQLLDVFGANIYLPVLSLDLTSQDDKGIALLQARATPRSLLSQVRLVVR 420
 QY 421 RALCOAGQQAQINQOLDIPPMILSYLKHQL 449
 DB 421 RALCOAGQQAQINQOLDIPPMILSYLKHQL 449
 RESULT 4
 AAU20555
 ID AAU20555 standard; Protein; 580 AA.
 AC AAU20555;
 XX 06-DEC-2001 (first entry)
 XX Human secreted protein, Seq ID No 547.
 XX Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
 KW rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
 KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
 KW cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
 KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
 KW Gaucher's disease; neurological disease; cerebrovascular disorder;
 KW thrombosis; wound healing.
 XX Homo sapiens.
 OS
 XX
 XX WO200155326-A2.
 XX 02-AUG-2001.
 XX 17-JAN-2001; 2001WO-US01347.
 XX 31-JAN-2000; 2000US-0179065.
 PR (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-451931/48.
 XX N-PSDB; AAS33264.
 DR
 XX New nucleic acids and polypeptides, useful for diagnosing, preventing
 PT or treating medical conditions -
 PT
 PS Claim 11; SEQ ID No 547; 753pp; English.
 XX The invention relates to novel isolated nucleic acid molecules (I)
 CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
 CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
 CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
 CC the prevention, treatment and diagnosis of diseases associated with
 CC inappropriate expression of secreted proteins. (I) and complementary
 CC sequences may also be used as DNA probes in diagnostic assays (e.g.
 CC polymerase chain reactions (PCR)) to detect and quantitate the presence
 CC of similar nucleic acid sequences in samples, and so which patients may
 CC be in need of restorative therapy. (II) may also be used as antigens in
 CC the production of antibodies and in assays to identify modulators
 CC (agonists and antagonists) of the expression and activity of the secreted
 CC proteins. The anti-(I) antibodies and antagonists may also be used to
 CC down regulate expression and activity of (II). The anti-(II) antibodies
 CC may also be used as diagnostic agents for detecting the presence of (II)

CC in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
 CC disorders include for example: immune/autoimmune diseases (e.g. HIV
 CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
 CC Caucher's disease), neurological diseases (e.g. Alzheimer's disease,
 CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
 CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
 CC angina and thrombosis), infections caused by bacteria, viruses and
 CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
 CC agonists, antagonists and antibodies can also be used to promote wound
 CC healing, maintain tissues before transplantation, and support cell culture
 CC of primary tissues. AAU20342-AAU20666 represent human secreted protein
 CC amino acid sequences, and related sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO
 CC at: ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 580 AA;

Query Match 99.5%; Score 2331; DB 22; Length 580;
 Best Local Similarity 99.6%; Pred. No. 8.7e-251;
 Matches 447; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MRESGWKLIDPISDFGRMGIPNPNWTITDANRNYEICSTYPPPEIVVPKSVTLGTVVGSSK 60
 DB 132 MRESGWKLIDPISDFGRMGIPNPNWTITDANRNYEICSTYPPPEIVVPKSVTLGTVVGSSK 191
 QY 61 FRSKERPVPVLSYLYKKNNAACRCQPLSGFYTRCVDDLELLEAISQTNPGSQFMYVVDV 120
 DB 192 FRSKERPVPVLSYLYKKNNAACRCQPLSGFYTRCVDDLELLEAISQTNPGSQFMYVVDV 251
 QY 121 RKIWHFLVIMRIVLQAKNLMIDITKIFSLQDPKKEEDTDEKQALNAQVYDNDY 180
 DB 252 RKIWHFLVIMRIVLQAKNLMIDITKIFSLQDPKKEEDTDEKQALNAQVYDNDY 311
 QY 181 TLDQLLRQERYKRFINSRSGWPGVPTPLRLAASYGHLSCLOVLLAHGADVDSLQVKAQTP 240
 DB 312 TLDQLLRQERYKRFINSRSGWPGVPTPLRLAASYGHLSCLOVLLAHGADVDSLQVKAQTP 371
 QY 241 LFTAVSHGHLDVRLVLEAGSPGSIYNNCSFVLTAARDGAVAILQELLDHGAENVKA 300
 DB 372 LFTAVSHGHLDVRLVLEAGSPGSIYNNCSFVLTAARDGAVAILQELLDHGAENVKA 431
 QY 301 KLPVWASNIASCSGPLYLAAYVGHLDVRLVLEAGSPGSIYNNCSFVLTAARDGAVAILQELLDHGAENVKA 431
 DB 432 KLPVWASNIASCSGPLYLAAYVGHLDVRLVLEAGSPGSIYNNCSFVLTAARDGAVAILQELLDHGAENVKA 491
 QY 361 CLHNCPEPIYQLLDVFGANIYLPVLSLDLTSQDDKGIALLQARATPRSLLSQVRLVVR 420
 DB 492 CLHNCPEPIYQLLDVFGANIYLPVLSLDLTSQDDKGIALLQARATPRSLLSQVRLVVR 551
 QY 421 RALCOAGQQAQINQOLDIPPMILSYLKHQL 449
 DB 552 RALCOAGQQAQINQOLDIPPMILSYLKHQL 580

RESULT 5
 AAU77411
 ID AAU77411 standard; Protein; 360 AA.
 XX
 AC AAU77411;
 XX 05-JUN-2002 (first entry)
 XX Human NOV5b protein, homologue of ankyrin-repeat proteins.
 XX Human; NOV5b-associated disorder; developmental disorder; blood disorder;
 KW endocrine disorder; vascular disease; gastrointestinal disorder; cancer;
 KW respiratory disorder; inflammatory disorder; reproductive disorder;
 KW neurodegenerative disorder; autoimmune disorder; infectious disease;
 KW cardiovascular disorder; cell signal processing; ankyrin-repeat protein;
 KW metabolic pathway modulation; NOV5b.

OS Homo sapiens.
PN WO200206329-A2.
XX 24-JAN-2002.
XX 18-JUL-2001; 2001WO-US22709.
XX 18-JUL-2000; 2000US-218870P.
XX 18-JUL-2000; 2000US-218875P.
XX 18-JUL-2000; 2000US-218901P.
XX 24-JUL-2000; 2000US-220273P.
XX 26-JUL-2000; 2000US-220912P.
XX 27-JUL-2000; 2000US-221233P.
XX 28-JUL-2000; 2000US-221650P.
XX (CURA-) CURAGEN CORP.
XX Rastelli L, Shimkets RA, Zerhusen B, Malyankar UM, Padigar M;
XX WPI: 2002-179781/23.
XX N-PSDB; ABK1107.
XX Novel cytoplasmic, nuclear, membrane bound and secreted NOVX
PT polypeptides, useful for treating developmental disorders, endocrine
PT disorders, vascular disorders, infectious diseases and
PT neurodegenerative disorders -
XX Claim 1; Page 40; 178pp; English.
XX The present invention relates to the isolation of novel human
CC polypeptides referred to as NOV1, NOV2, NOV3, NOV4a, NOV4b, NOV5a,
CC NOV5b, NOV6 AND NOV7, and the polynucleotide sequences encoding them.
CC The NOVX polypeptides are related to NOPE, cadherin, interferon
CC alpha-13, ADAM, ankyrin repeat-containing, transpanin or semaphorin
CC polypeptides. The sequences of the invention are useful for identifying
CC an agent (a cellular receptor or downstream effector) that binds to
CC the NOVX polypeptide, or an agent that modulates its expression or
CC activity. They are useful for treating or preventing NOVX-associated
CC disorders such as developmental disorders, endocrine disorders, vascular
CC diseases, gastrointestinal disorders, respiratory disorders, inflammatory
CC disorders, blood disorders, reproductive disorders, neurodegenerative
CC disorders, autoimmune and immune disorders, infectious diseases,
CC cardiovascular disorders, cancers, and other disorders related to cell
CC signal processing and metabolic pathway modulation. The present sequence
CC represents the human NOV5b protein.
XX Sequence 360 AA;
SQ Query Match 72.1%; Score 1690; DB 23; Length 360;
Best Local Similarity 99.4%; Pred. No. 1.7e-179;
Matches 326; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 122 PKIWHFLVLMRIVLQAKNLMIDITKIFSLQPKDEEDTDEEKQALNQAVYNDST 181
DB 33 PQIWHFLVLMRIVLQAKNLMIDITKIFSLQPKDEEDTDEEKQALNQAVYNDST 92
QY 182 LQQLRQERYKRFINSRSGVPGTPLRLAASVGHLSCLQVLLAHGADVSDLVKATPL 241
DB 93 LQQLRQERYKRFINSRSGVPGTPLRLAASVGHLSCLQVLLAHGADVSDLVKATPL 152
QY 242 FTAVSHGHLDVVRVLEAGASPGSYNNCSPLVTAARGAVAILQELLDHGAENVKAK 301
DB 153 FTAVSHGHLDVVRVLEAGASPGSYNNCSPLVTAARDSAVAILQELLDHGAENVKAK 212
QY 302 LPWASNIASCSPGLYLAAYVGHLDVVRVLEAGADPDYNDCTDQGLARVPRRTLEIC 361
DB 213 LPWASNIASCSPGLYLAAYVGHLDVVRVLEAGADPDYNDCTDQGLARVPRRTLEIC 272
QY 362 LHHNCEPEYIQLIDFGANIYLPFSLDLTSDQDKGIALLLQARATPRSLLSQVRLVRR 421
DB 273 LHHNCEPEYIQLIDFGANIYLPFSLDLTSDQDKGIALLLQARATPRSLLSQVRLVRR 332
SQ

QY 422 ALCQAGQPOAINQLDIPPMILSYLKHQL 449
DB 333 ALCQAGQPOAINQLDIPPMILSYLKHQL 360
RESULT 6
ABP69316
ID ABP69316 standard; Protein; 318 AA.
XX AC ABP69316;
XX 20-JAN-2003 (first entry)
XX Human polypeptide SEQ ID NO 1363.
XX Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW cell-proliferative disorder; neurodegenerative disease; bacterial;
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;
KW antiarthritic.
XX Homo sapiens.
OS WO200270539-A2.
XX 12-SEP-2002.
XX 05-MAR-2002; 2002WO-US05095.
XX 05-MAR-2001; 2001US-0799451.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
PI Wehrman T, Wang J, Wang D, Drmanac RT;
XX WPI: 2002-759812/82.
XX N-PSDB; AB211533.
XX New polynucleotides comprising sequences assembled from expressed
PT sequence tags (ESTs), useful for treating cell-proliferative,
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
PT platelet or coagulation disorders -
XX Claim 9; SEQ ID NO 1363; 1012pp + Sequence Listing; English.
XX The invention relates to an isolated polynucleotide (I) comprising a
CC nucleotide sequence selected from any of 948 sequences
CC (AB211119-AB212066) or their mature protein coding portion, active domain
CC coding protein or complementary sequences. The polynucleotides are useful
CC for identifying expressed genes or for physical mapping of human genome.
CC The encoded polypeptides (ABP6902-ABP6949) are useful as molecular
CC weight markers, as a food supplement, for generating antibodies, in
CC medical imaging, screening and diagnostic assays and for treating
CC cell-proliferative disorders (cancer), neurodegenerative diseases
CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
CC disorders, platelet or coagulation disorders, wound, burns, incision,
CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
CC parasitic), arthritis, etc.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 318 AA;
SQ Query Match 69.8%; Score 1635; DB 23; Length 318;
Best Local Similarity 99.4%; Pred. No. 2e-173;

Matches 316; Conservative 1; Mismatches 1; Indels 0; Gaps 0;	
QY 132	MRIVLQAKNLMIDITKIFSLLOPDKEEDTDTTEKQALNQAVYDNDSTYLDQLLRQERY 191
Db 1	MRIVLQAKNLMIDITKIFSLLOPDKEEDTDTTEKQALNQAVYDNDSTYLDQLLRQERY 60
QY 192	KRFINSRSGWGPVGTPLRLAASYGHLSCLOVLLAHGADVDSLQVKAQTPLFTAVSHGHLD 251
Db 61	KRFINSRSGWGPVGTPLRLAASYGHLSCLOVLLAHGADVDSLQVKAQTPLFTAVSHGHLD 120
QY 252	CVRVLEAGASPGSGIYNNCSPLVTAARDGAVAILQELLDHGAEANVKAKLPVWASNIAS 311
Db 121	CVRVLEAGASPGSGIYNNCSPLVTAARDGAVAILQELLDHGAEANVKAKLPVWASNIAS 180
QY 312	CSGPLYLAAYVGHLCDFRLLHGHADPDYNCCTDQGLLARVPRPTLLEICLHHNCEPEYI 371
Db 181	CSGPLYLAAYVGHLCDFRLLHGHADPDYNCCTDQGLLARVPRPTLLEICLHHNCEPEYI 240
QY 372	QLLIDFGANIYLPSSLDLTSQDDKGIALLLQARATPRSLLSQVRLVVRRALCOAQPOQA 431
Db 241	QLLIDFGANIYLPSSLDLTSQDDKGIALLLQARATPRSLLSQVRLVVRRALCOAQPOQA 300
QY 432	INQDIPPMILSYLKHQL 449
Db 301	INQDIPPMILSYLKHQL 318
RESULT 7	
AAU77410	AAU77410 standard; Protein; 351 AA.
XX AC	AAU77410;
XX DT	05-JUN-2002 (first entry)
XX DE	Human NOV5a protein, homologue of ankyrin-repeat proteins.
XX KW	Human; NOV5a-associated disorder; developmental disorder; blood disorder; endocrine disorder; vascular disease; gastrointestinal disorder; cancer; respiratory disorder; inflammatory disorder; reproductive disorder; neurodegenerative disorder; autoimmune disorder; infectious disease; cardiovascular disorder; cell signal processing; ankyrin-repeat protein; metabolic pathway modulation; NOV5a.
XX OS	Homo sapiens.
XX PN	WO200206329-A2.
XX PD	24-JAN-2002.
XX PF	18-JUL-2001; 2001WO-US22709.
XX PR	18-JUL-2000; 2000US-218870P.
XX PR	18-JUL-2000; 2000US-218875P.
XX PR	18-JUL-2000; 2000US-218901P.
XX PR	24-JUL-2000; 2000US-220273P.
XX PR	26-JUL-2000; 2000US-220912P.
XX PR	27-JUL-2000; 2000US-221233P.
XX PR	28-JUL-2000; 2000US-221650P.
XX PA	(CURA-) CURAGEN CORP.
XX PI	Rastelli L, Shimkets RA, Zerhusen B, Malyankar UM, Padigar M;
XX DR	WPI; 2002-179781/23.
XX DR	N-PSDB; ABK11106.
XX PT	Novel cytoplasmic, nuclear, membrane bound and secreted NOV5 polypeptides, useful for treating developmental disorders, endocrine disorders, vascular disorders, infectious diseases and neurodegenerative disorders -
XX PS	Claim 1; Page 39; 178pp; English.

XX	The present invention relates to the isolation of novel human polypeptides referred to as NOV1, NOV2, NOV3, NOV4a, NOV4b, NOV5a, NOV5b, NOV6 AND NOV7, and the polynucleotide sequences encoding them. The NOV5 polypeptides are related to NOV6, cadherin, interferon alpha-13, ADAM, ankyrin repeat-containing, transpanin or semaphorin polypeptides. The sequences of the invention are useful for identifying an agent (a cellular receptor or downstream effector) that binds to the NOV5 polypeptide, or an agent that modulates its expression or activity. They are useful for treating or preventing NOV5-associated disorders such as developmental disorders, endocrine disorders, vascular diseases, gastrointestinal disorders, respiratory disorders, inflammatory disorders, blood disorders, reproductive disorders, neurodegenerative disorders, autoimmune and immune disorders, infectious diseases, cardiovascular disorders, cancers, and other disorders related to cell signal processing and metabolic pathway modulation. The present sequence represents the human NOV5a protein.
XX SQ	Sequence 351 AA;
Query Match 69.5%; Score 1629; DB 23; Length 351;	
Best Local Similarity 97.5%; Pred. No. 1.1e-172;	
Matches 318; Conservative 0; Mismatches 0; Indels 8; Gaps 1;	
QY 132	MRIVLQAKNLMIDITKIFSLLOPDKEEDTDTTEKQALNQAVYDNDSTYLDQLLRQERY 191
Db 26	MRIVLQAKNLMIDITKIFSLLOPDKEEDTDTTEKQALNQAVYDNDSTYLDQLLRQERY 85
QY 192	KRFINSRSGWGPVGTPLRLAASYGHLSCLOVLLAHGADVDSLQVKAQTPLFTAVSHGHLD 251
Db 86	KRFINSRSGWGPVGTPLRLAASYGHLSCLOVLLAHGADVDSLQVKAQTPLFTAVSHGHLD 145
QY 252	CVRVLEAGASPGSGIYNNCSPLVTAARDGAVAILQELLDHGAEANVKAKLPVWASNIAS 311
Db 146	CVRVLEAGASPGSGIYNNCSPLVTAARDGAVAILQELLDHGAEANVKAKLPVWASNIAS 205
QY 312	CSGPLYLAAYVGHLCDFRLLHGHADPDYNCCTDQGLLARVPRPTLLEICLHHNCEPEYI 371
Db 206	CSGPLYLAAYVGHLCDFRLLHGHADPDYNCCTDQGLLARVPRPTLLEICLHHNCEPEYI 265
QY 372	QLLIDFGANIYLPSSLDLTSQDDKGIALLLQARATPRSLLSQVRLVVRRAL 423
Db 266	QLLIDFGANIYLPSSLDLTSQDDKGIALLLQARATPRSLLSQVRLVVRRAL 325
QY 424	COAQPOQAINQDIPPMILSYLKHQL 449
Db 326	COAQPOQAINQDIPPMILSYLKHQL 351
RESULT 8	
ABB97599	ABB97599 standard; Protein; 370 AA.
XX AC	ABB97599;
XX DT	27-JUN-2002 (first entry)
XX DE	Novel human protein SEQ ID NO: 867.
XX KW	Human; antianaemic; vulnery; antiinflammatory; immunomodulator; antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy; neuroprotective; antiparkinsonian; protein therapy; EST; expressed sequence tag.
XX OS	Homo sapiens.
XX PN	WO200222660-A2.
XX PD	21-MAR-2002.
XX PF	10-SEP-2001; 2001WO-US26015.
XX PR	11-SEP-2000; 2000US-0659671.

```
XX (HYSE-) HYSEQ INC.
XX PA
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX DR WPI; 2002-292408/33.
XX DR N-PSDB; ABN32785.
XX PT An isolated polynucleotide for treating diseases associated with its
XX encoded polypeptide such as cancer and multiple sclerosis -
XX PS Example 2; SEQ ID NO 867; 509pp; English.
XX CC The present invention provides the protein and coding sequences of 444
XX novel human proteins. These were isolated from expressed sequences tags
XX (ESTs). They can be used to stimulate cell growth, to regulate
XX haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
XX e.g. in burn treatment, to regulate the immune system e.g. to treat
XX multiple sclerosis, to regulate activin or inhibin e.g. to treat
XX infertility, to regulate haemostasis or thrombolysis e.g. to treat
XX stroke and cancer, to screen for drugs, to treat inflammatory conditions
XX e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
XX Parkinson's disease. The present sequence is a protein of the invention.
XX SQ Sequence 370 AA;
Query Match 27.4%; Score 642.5; DB 23; Length 370;
Best Local Similarity 53.9%; Pred. No. 2.2e-62;
Matches 145; Conservative 20; Mismatches 33; Indels 71; Gaps 8;
QY 1 MRESGKLDIPDSDFGRMGIPNRNWTITDANRNYEICSTYPPPIVVPKSVTLGTVVGSSK 60
DB 122 MRESGKLDIPDSDFGRMGIPNRNWTITDANRNYEICSTYPPPIVVPKSVTLGTVVGSSK 181
QY 61 FRSKERPVLVLYKENNAICRCSOPLSGFYTRCVDDLELLEAISQTPNGSQFYVVDVT 120
DB 182 FRSKERPVLVLYKENNAICRCSOPLSGFYTRCVDDLELLEAISQTPNGSQFYVVDVT 241
QY 121 RPKIWHFLVLIWRIVLQAKMIMDITKIFSLLOPKDEEDTTEKQALNQAVYDNDY 180
DB 242 RPK-----AVKVEKASVL-----VHCSDGW 261
QY 181 -----TLDQLLRQRYKRF-----INSRSGWGPCTPLRLAASVGHLSCLQVLLAHGAD 229
DB 262 DRTAQCVSASILLDPFYTFKGLMLIEKEMISMGH--KFSQRCGHL-----D 308
QY 230 VDSLDVKAQTPLTAVSHGHLDVCRVILE 258
DB 309 GDSKEV---SPIFTQ---FLDCIWQLME 330
RESULT 9
AAE32130
ID AAE32130 standard; Protein; 337 AA.
XX AC AAE32130;
XX DT 24-MAR-2003 (first entry)
DE Human cytoskeleton-associated protein, CSAP-28.
XX DE
XX KW Human cytoskeleton-associated protein; CSAP-28; atherosclerosis;
XX KW cancer; gene therapy.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Domain 245..270
XX FT /note= "Transmembrane domain"
XX PN WO200279404-A2.
XX
```

```
PD 10-OCT-2002.
XX 25-MAR-2002; 2002WO-US09288.
XX 29-MAR-2001; 2001US-280508P.
PR 03-APR-2001; 2001US-281323P.
PR 13-APR-2001; 2001US-283769P.
PR 04-MAY-2001; 2001US-288609P.
PR 10-MAY-2001; 2001US-290518P.
PR 18-MAY-2001; 2001US-291870P.
PR 29-MAY-2001; 2001US-294451P.
XX (INCY-) INCYTE GENOMICS INC.
XX PA Hafalia AJA, Tang TY, Yue H, Khan FA, Ison CH, Baughn MR;
PI Warren BA, Duggan BM, Thangavelu K, Honchell CD, Azimzai Y;
PI Elliott VS, Burford N, Ding L, Yue H, Becha S, Emerling BM;
PI Richardson TW, Lee SY, Bandman O, Lal PG, Lee S, Gietzen KJ;
PI Walla NK, Griffin JA, Lee EA, Swarnakar A, Ring HZ, Jones KA;
XX WPI: 2003-092894/08.
DR N-PSDB; AAD49617.
XX New human cytoskeleton-associated proteins, useful for preparing a
PT composition for diagnosing or treating a disease or condition
PT associated with decreased expression or overexpression of functional
PT CSAP e.g., cancer -
XX Claim 1; Page 202-203; 233pp; English.
PS The invention relates to new human cytoskeleton-associated protein
CC (CSAP) and its polynucleotide. The polypeptide is useful for preparing
CC a composition for diagnosing or treating a disease or condition
CC associated with decreased expression or overexpression of functional
CC CSAP e.g. atherosclerosis or cancer. The present sequence is human
CC CSAP-28 protein. The invention is useful in gene therapy.
XX SQ Sequence 337 AA;
Query Match 21.3%; Score 499; DB 24; Length 337;
Best Local Similarity 39.1%; Pred. No. 2e-46;
Matches 119; Conservative 57; Mismatches 102; Indels 26; Gaps 9;
QY 166 EKOALNQAVYDNDYVTLQDLROERYKRFINSRSGW---GVPGTPLRLAASVGHLSCLQV 222
DB 39 EDTRLHDAAYVGDLOTLSLQESYRKRINEKSVCCGWLCTPLRIATAGHSCVDF 98
QY 223 LLAHGADVDSLVDVKAQTPLTAVSHGHLDVCRVILEAGASPGSGSYNNCSPLVTAARDGA 282
DB 99 LIRKGAEDVLDVKGQTALYVAVVNGHLESTQILLEAGADPNGRHRRSTPVVHASRVGR 158
QY 283 VAILQELLDHGAENNVKAKL-----PVWASNTAS---CSGPLYLAAYVGHLCDFRLLLLH 334
DB 159 ADILKALIRYGDVVDVNHHLTPDVQFRFSRRLTSLVWC--PLYISAAYHNLQCFRLLLLA 216
QY 335 GADPDYNC-----TDOGLLARVPRPRTLLEICLHNCPEYIOLLIDFGANIYL---PSLS 387
DB 217 GANPDFNCNGPVNTQGFYR--GSPCGVMDAVLRHGCCEAAVSLVFEFGANLNLVKWESLG 274
QY 388 LDLTQS---DDKGIALLQARATPRSLLSQVRLVVRRLCQAGQPOAINQLDIPPMILISY 444
DB 275 PESRRGRKVDPEALQVFKEARSVPTLLCLCHVAVRRLGK-HRLHLIFSLPLPDPIKPF 333
QY 445 LKHQ 448
DB 334 LLHE 337
RESULT 10
AAU20362
ID AAU20362 standard; Protein; 341 AA.
XX AC AAU20362;
```

XX DT 04-DEC-2001 (first entry)

XX DE Human secreted protein, Seq ID No 354.

XX KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina; rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;

XX KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;

XX KW cytotatic; Alzheimer's disease; Parkinson's disease; human; cancer;

XX KW multiple sclerosis; cancer; hyperproliferative disorder; infection;

XX KW Gaucher's disease; neurological disease; cerebrovascular disorder;

XX KW thrombosis; wound healing.

XX OS Homo sapiens.

XX PN WO200155326-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01347.

XX PR 31-JAN-2000; 2000US-0179065.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Barash SC, Ruben SM;

XX DR WPI; 2001-451931/48.

XX DR N-PSDB; AAS33071.

XX PT New nucleic acids and polypeptides, useful for diagnosing, preventing or treating medical conditions -

XX PS Claim 11; SEQ ID No 354; 753pp; English.

XX CC The invention relates to novel isolated nucleic acid molecules (I) encoding human secreted proteins (II). (I) and (II) are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate expression of secreted proteins. (I) and complementary sequences may also be used as DNA probes in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate the presence of similar nucleic acid sequences in samples, and so which patients may be in need of restorative therapy. (II) may also be used as antigens in the production of antibodies and in assays to identify modulators (agonists and antagonists) of the expression and activity of the secreted proteins. The anti-(II) antibodies and antagonists may also be used to down regulate expression and activity of (II). The anti-(II) antibodies may also be used as diagnostic agents for detecting the presence of (II) in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The disorders include for example: immune/autoimmune diseases (e.g. HIV (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis and multiple sclerosis), cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of the breast or liver, Sezary syndrome and Gaucher's disease), neurological diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/cerebrovascular disorders (e.g. cardiac arrest, tachycardia, angina and thrombosis), infections caused by bacteria, viruses and fungi and ocular disorders (e.g. corneal infections). (I) and (II), agonists, antagonists and antibodies can also be used to promote wound healing, maintain organs before transplantation, and support cell culture of primary tissues. AAU20342-AAU20666 represent human secreted protein amino acid sequences, and related sequences of the invention.

XX CC Note: The sequence data for this patent did not appear in the printed specification but was obtained in electronic format directly from WIPO at: ftp.wipo.int/pub/published_pct_sequences.

XX CC Sequence 341 AA;

XX SQ Query Match 21.3%; Score 499; DB 22; Length 341; Best Local Similarity 39.1%; Pred. No. 2.1e-46; Matches 119; Conservative 57; Mismatches 102; Indels 26; Gaps 9;

QY 166 EQALNQAVYDNDVSYTLQDLRQERYKRFINSRGW---GVFCTPLRLAASYCHLSCLQV 222

DB 43 EDTRLDAAYVGDQLQTLRSLLQEEYSRPRINEKSVWCCGWLCTPLRIATAAGHSCVDF 102

QY 223 LLAHGADVSLDVKAQTPLFTAVSHGHLDVCRVLLLEAGASPGSGSYNNCSPLVLTAAADGA 282

DB 103 LIRKGAEDVLVDKGTALYVAVVNGHLESTQILLEAGADPNRSHRSTPVPVHASRVGR 162

QY 283 VAILQELLDHGAENVKAKL-----PVWASNTAS---CSGPLYLAAYVGHLCDFRLLLLH 334

DB 163 ADILKALIRYGVADVDVNHHLTPDVQFRFRRLTSLVVC--PLYISAAYHNLQCFRLLLA 220

QY 335 GADPDVNC---TDOGLLARVPRPTLLLEICLHNCEPEYIQLLIDFGANIVL---PSLS 387

DB 221 GANPFCNCPVNTQGYR--GSPGCVMDVLRHCEAFVSLVFEFGANLNLVKWESLG 278

QY 388 LDLTSG---DDKGIALLQARATPRSLISQVRLVRRALCQAGQPOAINQLDIPPMLSY 444

DB 279 PESGRKRVDPALQVFKEARSPRTLCLCRVAVRRALGK-HRLHLPSLPDPDIKKF 337

QY 445 LKHQ 448

DB 338 LLHE 341

RESULT 11

AAU20580

ID AAU20580 standard; Protein; 343 AA.

XX AC AAU20580;

XX DT 04-DEC-2001 (first entry)

XX DE Human secreted protein, Seq ID No 572.

XX KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina; rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;

XX KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;

XX KW cytotatic; Alzheimer's disease; Parkinson's disease; human; cancer;

XX KW multiple sclerosis; cancer; hyperproliferative disorder; infection;

XX KW Gaucher's disease; neurological disease; cerebrovascular disorder;

XX KW thrombosis; wound healing.

XX OS Homo sapiens.

XX PN WO200155326-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01347.

XX PR 31-JAN-2000; 2000US-0179065.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Barash SC, Ruben SM;

XX DR WPI; 2001-451931/48.

XX DR N-PSDB; AAS33289.

XX PT New nucleic acids and polypeptides, useful for diagnosing, preventing or treating medical conditions -

XX PS Claim 11; SEQ ID No 572; 753pp; English.

XX CC The invention relates to novel isolated nucleic acid molecules (I) encoding human secreted proteins (II). (I) and (II) are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate expression of secreted proteins. (I) and complementary sequences may also be used as DNA probes in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate the presence of similar nucleic acid sequences in samples, and so which patients may be in need of restorative therapy. (II) may also be used as antigens in the production of antibodies and in assays to identify modulators (agonists and antagonists) of the expression and activity of the secreted proteins. The anti-(II) antibodies and antagonists may also be used to down regulate expression and activity of (II). The anti-(II) antibodies may also be used as diagnostic agents for detecting the presence of (II) in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The disorders include for example: immune/autoimmune diseases (e.g. HIV (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis and multiple sclerosis), cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of the breast or liver, Sezary syndrome and Gaucher's disease), neurological diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/cerebrovascular disorders (e.g. cardiac arrest, tachycardia, angina and thrombosis), infections caused by bacteria, viruses and fungi and ocular disorders (e.g. corneal infections). (I) and (II), agonists, antagonists and antibodies can also be used to promote wound healing, maintain organs before transplantation, and support cell culture of primary tissues. AAU20342-AAU20666 represent human secreted protein amino acid sequences, and related sequences of the invention.

XX CC Note: The sequence data for this patent did not appear in the printed specification but was obtained in electronic format directly from WIPO at: ftp.wipo.int/pub/published_pct_sequences.

XX CC Sequence 341 AA;

XX SQ Query Match 21.3%; Score 499; DB 22; Length 341; Best Local Similarity 39.1%; Pred. No. 2.1e-46; Matches 119; Conservative 57; Mismatches 102; Indels 26; Gaps 9;

CC polymerase chain reactions (PCR)) to detect and quantitate the presence
 CC of similar nucleic acid sequences in samples, and so which patients may
 CC be in need of restorative therapy. (II) may also be used as antigens in
 CC the production of antibodies and in assays to identify modulators
 CC (agonists and antagonists) of the expression and activity of the secreted
 CC proteins. The anti-(II) antibodies and antagonists may also be used to
 CC down regulate expression and activity of (II). The anti-(II) antibodies
 CC may also be used as diagnostic agents for detecting the presence of (II)
 CC in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The
 CC disorders include for example: immune/autoimmune diseases (e.g. HIV
 CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
 CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
 CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
 CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
 CC angina and thrombosis), infections caused by bacteria, viruses and
 CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
 CC agonists, antagonists and antibodies can also be used to promote wound
 CC healing, maintain organs before transplantation, and support cell culture
 CC of primary tissues. AAU20342-AAU20666 represent human secreted protein
 CC amino acid sequences, and related sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO
 CC at: ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 343 AA;

Query Match 21.3%; Score 499; DB 22; Length 343;
 Best Local Similarity 39.1%; Pred. No. 2.1e-46;
 Matches 119; Conservative 57; Mismatches 102; Indels 26; Gaps 9;
 QY 166 EKOALNQAVYDNDSTYLDQLLRQERYKRFINSRSGW---GVPTPLRLAASYGHLSCLOV 222
 Db 45 ETRLHDAAYVGDQLQTLRLSQEESYRSINEKSVCCWGLPCTPLRIATAGHGCVDV 104
 QY 223 LLAAGADVSLDKACTPLFTAVSHGHLCVRLLEAGASPGGSIYNNCSPLVLTAAADGA 282
 Db 105 LIRKGAEDLVVKGQALYVAVVNGHLESTQILLEAGADPNGRHRSPTVTHARVGR 164
 QY 283 VAILQELDHGAENYKAKL-----PWASNIAS---CSGPLYLAAYVGHLCDFRLLH 334
 Db 165 ADILKALIRYADVDVNHHLTPDVPFRSRLTSLVVC--PLYISAAYHNLQCFRLLILA 222
 QY 335 GADPDYNC---TDQGLLARVPRTLLLEICLHNCEPEYIQLIDFGANIYL---PSLS 387
 Db 223 GANPDNCGPVNTQGYR--GSPGCVMDAHLRGCEAAAFVSLVVEFGANLNLVKWESLG 280
 QY 388 LDITSQ---DDKGIALLQARATPRSLLSQVRLVVRRLCAQGPQAINQLDIPPMILISY 444
 Db 281 PESRGRKVDPEALQVFKEARSPVTRLLCLCRVAVRRALGK-HRLHLISPLPDPPIK 339
 QY 445 LKHQ 448
 Db 340 LLHE 343

RESULT 12
 AAU62621
 ID AAU62621 standard; Protein; 350 AA.

XX AAU62621;
 XX 27-OCT-1998 (first entry)

XX Mus musculus SOCS7 protein.

XX SOCS; suppressor of cytokine signalling; PCR primer;
 XX autoimmune disease; diagnosis; cancer; treatment;
 XX cytokine mediated cellular responsiveness; hyperimmunity;
 XX immunosuppression; allergies; hypertension.
 XX Mus musculus.

XX Key Location/Qualifiers
 FH Misc-difference 167
 FT /note= "undefined amino acid"

XX WO9820023-A1.

XX 14-MAY-1998.

XX 31-OCT-1997; 97WO-AU00729.

XX 14-FEB-1997; 97AU-0005117.

XX 01-NOV-1996; 96AU-0003384.

XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.

XX Alexander WS, Hilton DJ, Metcalf D, Nicholson SE;

XX Nicola NA, Richardson RT, Starr R, Viney EM, Willson TA;

XX WPI; 1998-286854/25.

XX N-PSDB; AAV38672.

XX Suppressor of cytokine signalling proteins - useful to treat

XX Claim 13; Page 144-146; 325pp; English.

XX The sequence is that of a suppressor of cytokine signalling
 CC protein (SOCS). SOCS can be used to screen for naturally
 CC occurring antibodies to SOCS, which may occur, e.g. in some autoimmune
 CC diseases. Alternatively, specific antibodies can be used to
 CC screen for SOCS, which is useful as a knowledge of SOCS levels
 CC may be important for the diagnosis of certain cancers. Soluble
 CC SOCS polypeptides can be used to treat disease, injury or
 CC abnormality involving cytokine mediated cellular responsiveness,
 CC e.g. hyperimmunity, immunosuppression, allergies and hypertension.

XX Sequence 350 AA;

Query Match 20.5%; Score 480; DB 19; Length 350;
 Best Local Similarity 38.3%; Pred. No. 2.9e-44;
 Matches 114; Conservative 58; Mismatches 104; Indels 22; Gaps 8;

QY 170 LNOAVYDNDSTYLDQLLRQERYKRFINSRSGW---GVPTPLRLAASYGHLSCLOVLLAH 226

Db 56 LHDAAYVGDQLQTLRLSQEESYRSINEKSVCCWGLPCTPLRIATAGHGCVDVFLIRK 115

QY 227 GADVSLDKACTPLFTAVSHGHLCVRLLEAGASPGGSIYNNCSPLVLTAAADGAVAL 286

Db 116 GAEDVLVDVKGQALYVAVVNGHLESTQILLEAGADPNGRHRSPTVTHARVGRDDL 175

QY 287 QELLDHGAENYKAKL-----PWASNIAS---CSGPLYLAAYVGHLCDFRLLH 338

Db 176 KALIRYADVDVNHHLNSDTRPPFSRLTSLVVC--PLYISAAYHNLQCFRLLQAGANP 233

QY 339 DYNCTDQGLLARVP--PRTLLEICLHNCEPEYIQLIDFGANIYL---PSLSLDLTSQ 393

Db 234 DFNCNGPVNTQGYRSGPCVMDAHLRGCEAAAFVSLVVEFGANLNLVKWESLGPEARG 293

QY 394 ---DDKGIALLQARATPRSLLSQVRLVVRRLCAQGPQAINQLDIPPMILSYLKHQ 448

Db 294 RKMDPEALQVFKEARSPVTRLLSLCRVAVRRALGKY-RULHLVPSPLPDPPIKFFLLYE 350

RESULT 13

AAU73229

ID AAU73229 standard; Protein; 629 AA.

XX AAU73229;

XX 11-MAY-2001 (first entry)

XX Human phosphatase MTMR7 h.
DE Phosphatase; gene therapy; enzyme; cancer; pathophysiological hypoxia;
XX cardiac dysfunction; vascular disorder; myopathy; ectodermal dysplasia;
KW congenital muscle disorder; Papillon-Lefevre syndrome; Cowden disease;
KW Moebius syndrome; Bjornstad syndrome; Bannayan Zonana syndrome;
KW schizoprenia; hamartoma.
XX
OS Homo sapiens.
XX
PN WO200112819-A2.
XX
PD 22-FEB-2001.
XX
PF 11-AUG-2000; 2000WO-US22158.
XX
PR 13-AUG-1999; 99US-0149005.
XX
PA (SUGEN-) SUGEN INC.
XX
PI Plowman GD, Martinez R, Whyte D, Hill R, Flanagan P, Lioubin M;
XX
PI WPI; 2001-211226/21.
DR N-PSDB; AAF63581.
XX
XX New protein phosphatase polypeptide for diagnosing and treating
PT phosphatase related disorders such as cancer, schizoprenia, cardiac
PT dysfunction and/or vascular disorders
XX
PS Claim 6; Fig 5; 138pp; English.
XX
CC The present invention relates to phosphatase proteins and coding
CC sequences. The present sequence is one such phosphatase. Phosphatases are
CC enzymes that catalyse the dephosphorylation of proteins modified by
CC phosphorylation of serine, threonine or tyrosine residues. The
CC phosphatases are useful for treating a variety of diseases: for example
CC cancer e.g. breast, urogenital, prostate, head, neck, lung cancers,
CC synovial sarcomas, renal cell carcinoma, non-small cell lung cancer,
CC hepatocellular carcinoma, pancreatic endocrine tumours, stomach cancer,
CC glioblastoma, colorectal cancer and thyroid cancer, pathophysiological
CC hypoxia, cardiac dysfunction and/or vascular disorders, myopathies,
CC congenital muscle disorders, Papillon-Lefevre syndrome, Cowden disease,
CC ectodermal dysplasia, Moebius syndrome, Bjornstad syndrome, Bannayan
CC Zonana syndrome, schizoprenia and hamartomas.
XX
SQ Sequence 629 AA;
Query Match 18.1%; Score 423; DB 22; Length 629;
Best Local Similarity 60.2%; Pred. No. 1.8e-37;
Matches 74; Conservative 21; Mismatches 28; Indels 0; Gaps 0;
QY 2 RESGWLIDPISDFGRMGIPNRNWTITDANRNYEICSTYPPETVVPKSVTLGTVGSSKF 61
DB 197 REQQWLIDLSEYTRMGLPNHYWQLSDVNRDVRVCDSTYPTLYPKSATAHIIVGSSKF 256
QY 62 RSKERPVLVLYKNNAAICRCSQPLSGFYTRCVDDLELLAISQTNPGSQPMYVVDTR 121
DB 257 RSRREFVLSYYKDNHASICRSSQPLSGFSARCLEQMLQAIRKANPGSDFVYVVDTR 316
QY 122 PKI 124
DB 317 PKL 319
RESULT 14
ID ABP62857
XX ABP62857 standard; Protein; 482 AA.
XX
AC ABP62857;
XX
XX 14-OCT-2002 (first entry)
XX

DE Human polypeptide SEQ ID NO 294.
XX
KW Human; dermatological; neuroprotective; nootropic; cancer;
KW antiparkinsonian; immunostimulant; cytostatic; immunosuppressive;
KW antidiabetic; antiallergic; gene therapy; wound healing; tissue repair;
KW burn; central nervous system disorder; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; immune disorder;
KW autoimmune disease; multiple sclerosis; diabetes; allergy.
XX
OS Homo sapiens.
XX
PN WO200218424-A2.
XX
PD 07-MAR-2002.
XX
PF 31-AUG-2001; 2001WO-US27093.
XX
PR 01-SEP-2000; 2000US-0654935.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Asundi V, Zhou P, Xue AJ, Ren F, Zhang J, Wang J;
PI Zhao QA, Wang D, Liu C, Drmanac RT, Wehrman T;
XX
DR WPI; 2002-583321/62.
DR N-PSDB; ABQ93336.
XX
XX New polynucleotide and polypeptides, useful for treatment and diagnosis
PT of Alzheimer's, Parkinson's, Huntington's, amyotrophic lateral
PT sclerosis, immune deficiencies, cancer, autoimmune disorders, multiple
PT sclerosis, diabetes and allergies
XX
PS Claim 20; SEQ ID NO 294; 284pp + Sequence Listing; English.
XX
CC The invention relates to an isolated polynucleotide (I) comprising one of
CC 245 sequences (ABQ93288-ABQ93532). Treating a condition comprising
CC administering to a mammalian subject a composition comprising the protein
CC (II) encoded by (I) (ABP62809-ABP63053) or an antibody (III) to (II).
CC (I), (II) and (III) are useful for diagnostic evaluation of disorders.
CC (I) is useful for gene therapy of diseases and (II) can be used for
CC therapeutic treatment. Diseases that may be treated include wound healing
CC and tissue repair, burns, central nervous system disorders (e.g.
CC Alzheimer's, Parkinson's, Huntington's and amyotrophic lateral
CC sclerosis), immune deficiencies, cancer, autoimmune disorders, multiple
CC sclerosis, diabetes and allergies.
CC Note: the sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 482 AA;
Query Match 16.8%; Score 392.5; DB 23; Length 482;
Best Local Similarity 44.6%; Pred. No. 2.9e-34;
Matches 74; Conservative 22; Mismatches 27; Indels 43; Gaps 1;
QY 2 RESGWLIDPISDFGRMGIPNRNWTITDANRNY----- 34
DB 7 REQQWLIDLSEYTRMGLPNHYWQLSDVNRDVRVCDSTYPTLYPKSATAHIIVGSSKFVLSYKNN 66
QY 35 -----EICSTYPPETVVPKSVTLGTVGSSKFSKERVPLVLSYKNN 78
DB 67 FSCGFSTPTTSSRVIKVCDSTYPTLYPKSATAHIIVGSSKFSRFRFPVLSYKDNH 126
QY 79 AAICRCSQPLSGFYTRCVDDLELLAISQTNPGSQPMYVVDTRPKI 124
DB 127 ASICRSSQPLSGFSARCLEQMLQAIRKANPGSDFVYVVDTRPKL 172
RESULT 15
ID ABB65465
XX ABB65465 standard; Protein; 639 AA.
XX
AC ABB65465;

XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 23187.
DE Drosophila melanogaster polypeptide SEQ ID NO 23187.
XX Drosophila melanogaster polypeptide SEQ ID NO 23187.
KW Drosophila melanogaster polypeptide SEQ ID NO 23187.
XX Drosophila melanogaster polypeptide SEQ ID NO 23187.
OS Drosophila melanogaster polypeptide SEQ ID NO 23187.
XX WO200171042-A2.
PN WO200171042-A2.
XX 27-SEP-2001.
PD 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
PF 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
PR 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
PA (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR WPI; 2001-656860/75.
XX N-PSDB; ABL09568.
DR N-PSDB; ABL09568.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Disclosure; SEQ ID NO 23187; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AB957737-AB972072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 639 AA;
SQ Sequence 639 AA;
Query Match 15.4%; Score 360; DB 22; Length 639;
Best Local Similarity 26.9%; Pred. No. 2e-30;
Matches 119; Conservative 62; Mismatches 159; Indels 102; Gaps 15;
QY 2 RESGKLIIDPISDFGRMGIPNRRWTITDANRYEIGCTYPPEIVVPKSVTLGTVVGSSKF 61
DB 71 KRGWDYFKLEAEFKMLVFNENATLCSNKEKELCDTPRQIYVPKREATMLLSSSRF 130
QY 62 RSKERVPLSYLYKENNAICRCQPLSGFYTRCVDLLELLEAISQTFNPGSQFMYYVDTR 121
DB 131 RSKGRPLVLYLH-NNKASICRCQPLSGFSARCLEDEQMLEAIRKTNSTNDYMYVDTR 189
QY 122 PKI-----WHFLVLMRIVLQAKNMLMDITKIFSLQPDKEE 159
DB 190 PRINAMANAAGKGYENAEFYENIKFHLG-ENIHVQRASLQ-----KVLEACEQK 240
QY 160 EDTDTEKQALN-----QAVYDNDSTYLDLQERYKRFINSRSGWVGP-----T 206
DB 241 SPTMSAFINALESSGWLKHSIRLDTSSFTANAV--DKGVSVVYVHCSDGMDRTAQVCSLA 298
QY 207 PLRLAASGYHLSCLOVL-----LAHGADV-----SLDVKAOTPLFTAVSHGLDCV-- 253
DB 299 QLMNLYRTIKGFQALIEKDWLAFGHKFSERCHIQTDAREVSPFTQ-----FLDCTWQ 354
QY 254 -----RVLL-----EAGSPGGSIVNNCSFVLTAARDGAVAILQELLHDGAEA 296
DB 355 LMSQRSEAFEFNRFLLILHDVHVSQCFGTGVCNE-----KORLDLKLAE 400

QY 297 NVKAKLPVWASNIASCSGPLYL-----AAVYGHLDGFRLLLLHGADPDVYNTDQGLIARVP 352
DB 401 RTFSLWGYMANHLNEYINPLYKPNVDEAIKANLAPQCICKFWEGM---YSRPFSGIHPREP 457
QY 353 RPRTLLEICLHNCBPEYIQLL 374
DB 458 LGDVLLDSKEHCNLSLEHDHVQHL 479

Search completed: December 1, 2003, 07:37:11
Job time : 72 secs

THIS PAGE BLANK (USPTO)


```
RESULT 5
US-09-155-078-2
; Sequence 2, Application US/09155078
; Patent No. 6312688
; GENERAL INFORMATION:
; APPLICANT: POUSTKA, Annemarie
; APPLICANT: KIOSCHIS, Petra
; APPLICANT: LAPORTE, Jocelyn
; APPLICANT: HU, Ling Jia
; APPLICANT: MANDEL, Jean Louis
; APPLICANT: DAHL, Niklas
; TITLE OF INVENTION: TYROSINE-PHOSPHATASE-RELATED PROTEIN
; FILE REFERENCE: 012627-010
; CURRENT APPLICATION NUMBER: US/09/155,078
; CURRENT FILING DATE: 1999-03-02
; EARLIER APPLICATION NUMBER: PCT/DE97/00592
; EARLIER FILING DATE: 1997-03-21
; EARLIER APPLICATION NUMBER: DE 196 11 234.6
; EARLIER FILING DATE: 1996-03-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 621
; TYPE: PRT
; ORGANISM: German-type microorganism & cell cul.
US-09-155-078-2

Query Match
Best Local Similarity 92.7% Score 264; DB 4; Length 621;
Matches 80; Conservative 46; Mismatches 106; Indels 58; Gaps 12;

QY 5 GWKLDIPDFGMRGIPNRNMTITDANRYEICSTYPPVVPKSVTLGTVVGSKFRSK 64
DB 181 GWTVPVVEYRROGLPNHNRITTEINKCYELCDTYPALLVVPYRASDDDLRRVATFRS 240
QY 65 ERVPVLSVLYKNNAAICRCSOPLSGFY-TRCVDDLELLEAISQNPQSFYVVDTRPK 123
DB 241 NRIPVLSWIHPENKTIVRCSOPLVMSGKRNKDEKYLVDVIRETNKQISKLTIIDARPS 300
QY 124 I-----WHFLVLMRIVLQAKNLM--DITKIFSLQDPKES----- 159
DB 301 VNAVANKATGGYESSDDAYHNAELFF---LDHNIHVRESLKKVKVDIVPVEESHWS 357
QY 160 --EDTDTEE--KQALNQAVYDNDSTYLDQLLRQERYKRFINSRSGVPGTFLRLA---- 211
DB 358 SLESTHWHLEHIKVLGTGAIQVADK-----VSSGKSSVLVHSCDGDWDRTAQLTSLAML 411
QY 212 -----ASYGHLSCLOVLLAHGADVSLDVKVQATPLFTAVSHGHLCV 253
DB 412 DSFYRSIEGFELTVQKEWISFGHKFASRI--GHG-DKNHTDAD-RSIPFLQ----FIDCV 463

RESULT 6
US-09-350-982C-5
; Sequence 5, Application US/09350982C
; Patent No. 6455290
; GENERAL INFORMATION:
; APPLICANT: Berthelsen, Jens
; APPLICANT: Toma, Salvatore
; APPLICANT: Isacchi, Antonella
; TITLE OF INVENTION: Tankyrase Homolog Protein(THP), Nucleic Acids, and Methods Rela
; TITLE OF INVENTION: Same
; FILE REFERENCE: PHRM-0043
; CURRENT APPLICATION NUMBER: US/09/350,982C
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 5
; LENGTH: 1166
; TYPE: PRT
; ORGANISM: Artificial
```

```
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (1102)..(1102)
OTHER INFORMATION: n is any nucleic acid
NAME/KEY: misc feature
LOCATION: (2650)..(2650)
OTHER INFORMATION: n is any nucleic acid
US-09-350-982C-5

Query Match
Best Local Similarity 9.3% Score 219; DB 4; Length 1166;
Matches 74; Conservative 26; Mismatches 78; Indels 66; Gaps 7;

QY 189 ERVKRF-----INRSRGVPGTFLRLAASGYHLSCLQVLLAHGADVSLDVKVQATPLFT 243
DB 38 ERVKRLVTPEKVNSTRDTAGRKSTPLHLAAGFGKVDVVEYLLONGANVQARDGGLIPLHN 97
QY 244 AVSHGHLDVVRVLLLEAGASPGSGSYNNCSPLVTAAARDGAVAILQELLDHGAENVKAKLP 303
DB 98 ACSFGHAEVNVNLLRHGADPNARDNWNYPFLHEAAIKGKIDVCIVLLQHGAEPTIRNTDG 157
QY 304 VWASNIA-----SC-----SGPYL 318
DB 158 RTALDLADPSAKAVLTGCKYKDELLESARSNGEKKWALLTPLNVNCHASDGRKSTPLHL 217
QY 319 AAVYGHLDLFRLLHGHADPDYNTDQGLLARVPRPRTLEICLHNC---EPEYIQLLI 375
DB 218 AAGYNRVYKIVQLLQHGAA--DVHAKRKGDL--VP-----LHNACSYGHYEVTLLV 264
QY 376 DFGA 379
DB 265 KHGA 268

RESULT 7
US-09-196-387-10
; Sequence 10, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
```

THIS PAGE BLANK (USPTO)

Db 347 GGGGCGGAGGTGGACCTGGTGGATGTCAGAGGGCAGAGCTCCCTGTATGTGGCTGTAGTG 406
Qy 1105 CATGGCCATCTGACCTGTGTACGTGTGCTTTTGGAGCTGGTGGCTCTCTGTGTGTAGC 1164
Db 407 AACGGGCACCTTGGAGACACTGAGATCTTTTGGAGCTGGTGTGATCCCAACGGCAGC 466
Qy 1165 ATCTACAAACACTGTTCTCCCGTCTCACAGCTGCCGCTGATGGTGTGTGTATCTCCTG 1224
Db 467 CGGCACACCGAGCACTCTGTGTACCATGCTTGTGTGGTAGGAGCAGCATCTCTG 526
Qy 1225 CAGAGCTCTAGACCATGTGTGAGGGCCAGCAAGTCAAGCTAAACT-----ACC 1274
Db 527 AAGGCTTATCAGGTATGGGCGAGATGTTGATGTCAACCATCATCTGAATTTCTGACACC 586
Qy 1275 AGTCTGGGCATCAACATAGCTTCATGTTCT-----GGCCCCCTCTATTTGGCGCA 1326
Db 587 CGGCCCCCTTTTACGGCGGCTAACCTCTTGTGTGTCTCTCTATACATCAGTGCT 646
Qy 1327 GTCTACGGGCACCTGGACTGTTTCCGCTGCTTTTGTCTCCAGGGGCGAGACCCTGACTAC 1386
Db 647 GCCTACCATCACTTCACTGCTTCAAGCTGCTTTCAGGCTGGGCGCAATCTGACTTC 706
Qy 1387 AACTGCACTGACAGGCG 1404
Db 707 AATTGCAATGGCCCTGTC 724

RESULT 2

US-09-302-769-26
; Sequence 26, Application US/09302769
; Patent No. 6323317
; GENERAL INFORMATION:
; APPLICANT: HILTON, Douglas J
; APPLICANT: ALEXANDER, Warren S
; APPLICANT: VINEY, Elizabeth M
; APPLICANT: WILLSON, Tracey A
; APPLICANT: RICHARDSON, Rachael T
; APPLICANT: STARR, Robyn
; APPLICANT: NICHOLSON, Sandra E
; APPLICANT: METCALF, Donald
; APPLICANT: NICOLA, Nicos A
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
; FILE REFERENCE: 109762
; CURRENT APPLICATION NUMBER: US/09/302,769
; PRIOR FILING DATE: 1999-04-30
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 419
; TYPE: DNA
; ORGANISM: Human
US-09-302-769-26

Query Match 3.8%; Score 71.2; DB 4; Length 419;
Best Local Similarity 55.4%; Pred. No. 2.28-11;
Matches 164; Conservative 0; Mismatches 123; Indels 9; Gaps 1;
Qy 862 GAGAGCAGGCTCTCAATCAAGCAGTGTATGACAACTCTCTATCTTTGGACGAGCTT 921
Db 114 GAGACACAGGCTCCATGATGACGCTTACGTGGGACCTCCAGACCTCAGAGGCTA 173
Qy 922 TTGGCCAGGAGCGTTACAAAGTTTCAACAGCAGGAGTGCTCG-----GGT 972
Db 174 TTGAAGAGGAGAGCTACCGGAGCGCATCAACAGAGTCTGTGTGTGTGGCTGG 233
Qy 973 GTTCTGGGACACCTTGGCTGCTTCTTATGGCCACTTGAGCTGTGTTCAGATC 1032
Db 234 CTCCTCTGCACACCGTTGCGAATCGCGCCACTCGAGGCCATGAGGAGCTGTGTGACTTC 293
1033 CTCCTAGCCCATGCTGTGTGATGTTGACAGCTTGGATGTCAGGCGCACAGCGCCACTTTC 1092

Db 294 CTCATCCGAGAGGGGCGGAGGTGGATCTGGTGGAGCTAAAAAGGACAGACGGCCCTGTAT 353
Qy 1093 ACTGCTGTGACGTCATGGCCATCTGACTGTGTACGTGTGCTTTTGGAGCTGGTGC 1148
Db 354 GTGGCTGTGGTGAACGGGCACTAGAGAGTACCCAGATCTCTTCGAAAGCTGGCGC 409
RESULT 3
US-09-221-298-28
; Sequence 28, Application US/09221298
; Patent No. 6284241
; GENERAL INFORMATION:
; APPLICANT: XU, Jiangchun
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER
; FILE REFERENCE: 210121.471
; CURRENT APPLICATION NUMBER: US/09/221,298
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: modified_base
; LOCATION: (212)
; OTHER INFORMATION: Where n is a, c, g or t
US-09-221-298-28

Query Match 3.7%; Score 69.4; DB 3; Length 401;
Best Local Similarity 52.4%; Pred. No. 7.4e-11;
Matches 176; Conservative 0; Mismatches 157; Indels 3; Gaps 1;
Qy 360 AAAAGAGATGAGGGAAGTGGATGGAACCTGATTGACCAATATCAGACTTTGGGCGTAT 419
Db 52 AGAAAAATTTAACTGGATGGATGGACAGTTTACAAATCCAGTGAAGAATACAGGAGGCA 111
Qy 420 GGGATACCCACAGAACTGGACCAATACAGATGCCAAGAACTATGAGATATGCAG 479
Db 112 GGGCTTGGCCCAATCACCATTGGAGATAACTTTTATTAATAAGTGTCTATGAGCTTCGCA 171
Qy 480 CACTACCTCTCTGAAATAGTGTCTTAAATCTGTCTTACCTGGGAACGGTGTGGAAG 539
Db 172 CACTACCTCTCTTGTGTGTCTTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 231
Qy 540 TTCAAAGTTTCAAGATTAAGAACGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 599
Db 232 TGCAACTTTTAGTCCCGAAATCGAATTCAGTGTCTCATGGATTTCATCCAGAAATAA 291
Qy 600 TGCTGCCATTTGGCGCTGTAGCCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 656
Db 292 GACGGTCATTGTGCGTTGCAAGTCAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 351
Qy 657 TGATGAGCTCTTGTGGAGGCCATTAGCCAAACAAA 692
Db 352 TGATGAGAAATATCTCGATGTTATCAGGGAGACTAA 387

RESULT 4

US-09-155-078-1
; Sequence 1, Application US/09155078
; Patent No. 6312688
; GENERAL INFORMATION:
; APPLICANT: POUSTKA, Annemarie
; APPLICANT: KIOSCHIS, Petra
; APPLICANT: LAPORTE, Jocelyn
; APPLICANT: HU, Ling Jia
; APPLICANT: MANDEL, Jean Louis
; APPLICANT: DAHL, Niklas
; TITLE OF INVENTION: TYROSINE-PHOSPHATASE-RELATED PROTEIN
; FILE REFERENCE: 012627-010
; CURRENT APPLICATION NUMBER: US/09/155,078

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2003, 07:37:16 ; Search time 539 Seconds
(without alignments)
153.646 Million cell updates/sec

Title: US-09-941-831A-20
Perfect score: 2343
Sequence: 1 MRESCWKLDPTSDFGRMGI.....QAINQLDIPPLMLSYLKHQL 449

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 673684 seqs, 184443283 residues

Total number of hits satisfying chosen parameters: 673684

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09D_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US10F_PUBCOMB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US10G_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2343	100.0	449	10	US-09-941-831-20
2	2343	100.0	449	12	US-10-207-175-18
3	1690	72.1	360	10	US-09-908-193-14
4	1629	69.5	351	10	US-09-908-193-12
5	1435	61.2	308	10	US-09-908-193-36
6	499	21.3	335	10	US-09-908-193-37
7	481	20.5	336	10	US-09-908-193-38
8	480.5	20.5	327	10	US-09-908-193-39
9	480	20.5	350	10	US-09-908-805B-25
10	454	19.4	271	10	US-09-908-193-40
11	233	12.5	458	11	US-09-811-469-10
12	293	12.5	458	11	US-09-811-469-11
13	293	12.5	458	12	US-10-370-659-10
14	293	12.5	458	12	US-10-370-659-11
15	264	11.3	603	11	US-09-919-039-134

16	246	10.5	323	9	US-09-884-870-2	Sequence 2, Appli
17	223	9.5	1181	12	US-10-199-937-139	Sequence 139, App
18	217.5	9.3	1262	12	US-10-199-937-107	Sequence 107, App
19	217.5	9.3	1385	12	US-10-199-937-133	Sequence 133, App
20	215	9.2	522	12	US-10-199-937-99	Sequence 99, Appli
21	215	9.2	1074	9	US-09-509-196A-2	Sequence 2, Appli
22	215	9.2	1166	11	US-09-972-115A-6	Sequence 6, Appli
23	215	9.2	1166	12	US-10-199-937-135	Sequence 135, App
24	215	9.2	1166	15	US-10-163-587A-15	Sequence 15, Appli
25	215	9.2	1169	12	US-10-199-937-2	Sequence 2, Appli
26	215	9.2	1169	12	US-10-199-937-101	Sequence 101, App
27	215	9.2	1227	12	US-09-849-602-26	Sequence 26, Appli
28	214	9.1	756	12	US-10-199-937-91	Sequence 91, Appli
29	214	9.1	784	12	US-10-199-937-89	Sequence 89, Appli
30	209	8.9	949	9	US-09-841-835-10	Sequence 10, Appli
31	208.5	8.9	1327	9	US-09-841-835-2	Sequence 2, Appli
32	208.5	8.9	1327	11	US-09-972-115A-8	Sequence 8, Appli
33	208.5	8.9	1327	12	US-10-199-937-4	Sequence 4, Appli
34	205	8.7	1724	10	US-09-964-899-43	Sequence 43, Appli
35	203	8.7	1762	12	US-10-205-194-117	Sequence 117, App
36	197	8.4	1184	12	US-10-117-229-3	Sequence 3, Appli
37	197	8.4	1184	12	US-10-117-229-9	Sequence 9, Appli
38	197	8.4	1771	12	US-10-117-229-4	Sequence 4, Appli
39	197	8.4	1771	12	US-10-117-229-11	Sequence 11, Appli
40	196.5	8.4	1715	12	US-10-021-571-4	Sequence 4, Appli
41	196.5	8.4	1715	12	US-10-117-229-2	Sequence 2, Appli
42	195	8.3	1619	12	US-10-369-978-4	Sequence 4, Appli
43	195	8.3	1704	12	US-10-369-978-2	Sequence 2, Appli
44	194.5	8.3	802	10	US-09-964-899-41	Sequence 41, Appli
45	194	8.3	1762	12	US-10-117-229-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-941-831-20
; Sequence 20, Application US/09941831
; Patent No. US20020160493A1
; GENERAL INFORMATION:
; APPLICANT: Ebner et al.
; TITLE OF INVENTION: P7049P1
; FILE REFERENCE: Serine/Threonine Phosphatase Polynucleotides, Polypeptides, and Antib
; CURRENT APPLICATION NUMBER: US/09/941,831
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: PCT/US01/06256
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/186,350
; PRIOR FILING DATE: 2000-03-02
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-941-831-20

Query Match 100.0%; Score 2343; DB 10; Length 449;
Best Local Similarity 100.0%; Pred. No. 4.2e-233;
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MRESCWKLDPTSDFGRMGIPNPNWTTIDANNNYICSTYPPPIVVPKSVTLGTVVGSSK	60
Db	1	MRESCWKLDPTSDFGRMGIPNPNWTTIDANNNYICSTYPPPIVVPKSVTLGTVVGSSK	60
Qy	61	FRSKERVPVLSYLYKENNAACRCSPQLSGFYTRCVDDELLEAISQTNPGSQFMVVDVT	120
Db	61	FRSKERVPVLSYLYKENNAACRCSPQLSGFYTRCVDDELLEAISQTNPGSQFMVVDVT	120
Qy	121	RPKIWHFLVLMIRIVLQAKNMLMDITIKFSLQDPKKEEDTDTTEKQALNQAQVYDNDSY	180
Db	121	RPKIWHFLVLMIRIVLQAKNMLMDITIKFSLQDPKKEEDTDTTEKQALNQAQVYDNDSY	180

181 TLDQLLRQERYKRFINSRSGWVPGTFLRLAASGYHLSCLQVLLAHGADVDSLVDVKAQTP 240
181 TLDQLLRQERYKRFINSRSGWVPGTFLRLAASGYHLSCLQVLLAHGADVDSLVDVKAQTP 240
241 LFTAVSHGHLDVVRVLEAGSPGSIYNNCSVPLTAARDGAVAILQELLDHGAENAVKA 300
241 LFTAVSHGHLDVVRVLEAGSPGSIYNNCSVPLTAARDGAVAILQELLDHGAENAVKA 300
301 KLPVWASNTASCSPGLYLAAYVGHLCDFRLLLLHGHADPDYNCDDQGLLARVPRPTLLEI 360
301 KLPVWASNTASCSPGLYLAAYVGHLCDFRLLLLHGHADPDYNCDDQGLLARVPRPTLLEI 360
361 CLHNCPEYIQLLIDFGANIYLPSSLDTSDDDKGIALLQARATPRSLLSQVRLVVR 420
361 CLHNCPEYIQLLIDFGANIYLPSSLDTSDDDKGIALLQARATPRSLLSQVRLVVR 420
421 RALCOAQOQAINQLDIPPMILSYLKHOL 449
421 RALCOAQOQAINQLDIPPMILSYLKHOL 449

RESULT 2
US-10-207-175-18
; Sequence 18, Application US/10207175
; Publication No. US20030157508A1
; GENERAL INFORMATION:
; APPLICANT: Ebner et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ74P1
; CURRENT APPLICATION NUMBER: US/10/207,175
; CURRENT FILING DATE: 2002-07-30
; PRIOR APPLICATION NUMBER: PCT/US01/01395
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: US 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/186,350
; PRIOR FILING DATE: 2000-03-02
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 18
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-207-175-18

Query Match 100.0%; Score 2343; DB 12; Length 449;
Best Local Similarity 100.0%; Pred. No. 4.2e-233;
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MRESGKGLDIPISDFGRMGIPNRRWITTDANRNYEICSTYPPPIVVPKSVTLGTVVGSSK 60
1 MRESGKGLDIPISDFGRMGIPNRRWITTDANRNYEICSTYPPPIVVPKSVTLGTVVGSSK 60
61 FRSKERVPLSYLYKENNAICRCSOPLSGFYTRCVDDELLEAISQTPGSGFMVYVDT 120
61 FRSKERVPLSYLYKENNAICRCSOPLSGFYTRCVDDELLEAISQTPGSGFMVYVDT 120
121 RPKIMHFLVLMIRVILQAKNMMDITKIFSLQDPKKEEDTTEEKQALNQAVYNDYS 180
121 RPKIMHFLVLMIRVILQAKNMMDITKIFSLQDPKKEEDTTEEKQALNQAVYNDYS 180
181 TLDQLLRQERYKRFINSRSGWVPGTFLRLAASGYHLSCLQVLLAHGADVDSLVDVKAQTP 240
181 TLDQLLRQERYKRFINSRSGWVPGTFLRLAASGYHLSCLQVLLAHGADVDSLVDVKAQTP 240
241 LFTAVSHGHLDVVRVLEAGSPGSIYNNCSVPLTAARDGAVAILQELLDHGAENAVKA 300
241 LFTAVSHGHLDVVRVLEAGSPGSIYNNCSVPLTAARDGAVAILQELLDHGAENAVKA 300
301 KLPVWASNTASCSPGLYLAAYVGHLCDFRLLLLHGHADPDYNCDDQGLLARVPRPTLLEI 360

301 KLPVWASNTASCSPGLYLAAYVGHLCDFRLLLLHGHADPDYNCDDQGLLARVPRPTLLEI 360
361 CLHNCPEYIQLLIDFGANIYLPSSLDTSDDDKGIALLQARATPRSLLSQVRLVVR 420
361 CLHNCPEYIQLLIDFGANIYLPSSLDTSDDDKGIALLQARATPRSLLSQVRLVVR 420
421 RALCOAQOQAINQLDIPPMILSYLKHOL 449
421 RALCOAQOQAINQLDIPPMILSYLKHOL 449

RESULT 3
US-09-908-193-14
; Sequence 14, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHIMKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
; APPLICANT: PADIGARU, MURALIDHARA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 21402-062
; CURRENT APPLICATION NUMBER: US/09/908,193
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: 60/220,273
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/221,650
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/221,233
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/220,912
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/218,875
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,870
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,901
; PRIOR FILING DATE: 2000-07-18
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 14
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: NOVX polypeptide
; OTHER INFORMATION: with homology to ankryrin repeat-containing-
; OTHER INFORMATION: (ASB-1) like protein
US-09-908-193-14

Query Match 72.1%; Score 1690; DB 10; Length 360;
Best Local Similarity 99.4%; Pred. No. 9.4e-166;
Matches 326; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

122 PKIMHFLVLMIRVILQAKNMMDITKIFSLQDPKKEEDTTEEKQALNQAVYNDYS 181
122 PKIMHFLVLMIRVILQAKNMMDITKIFSLQDPKKEEDTTEEKQALNQAVYNDYS 181
33 PQIMHFLVLMIRVILQAKNMMDITKIFSLQDPKKEEDTTEEKQALNQAVYNDYS 92
33 PQIMHFLVLMIRVILQAKNMMDITKIFSLQDPKKEEDTTEEKQALNQAVYNDYS 92
182 LPQLLRQERYKRFINSRSGWVPGTFLRLAASGYHLSCLQVLLAHGADVDSLVDVKAQTP 241
182 LPQLLRQERYKRFINSRSGWVPGTFLRLAASGYHLSCLQVLLAHGADVDSLVDVKAQTP 241
93 LPQLLRQERYKRFINSRSGWVPGTFLRLAASGYHLSCLQVLLAHGADVDSLVDVKAQTP 152
93 LPQLLRQERYKRFINSRSGWVPGTFLRLAASGYHLSCLQVLLAHGADVDSLVDVKAQTP 152
242 FTAVSHGHLDVVRVLEAGSPGSIYNNCSVPLTAARDGAVAILQELLDHGAENAVKA 301
242 FTAVSHGHLDVVRVLEAGSPGSIYNNCSVPLTAARDGAVAILQELLDHGAENAVKA 301
153 FTAVSHGHLDVVRVLEAGSPGSIYNNCSVPLTAARDGAVAILQELLDHGAENAVKA 212
153 FTAVSHGHLDVVRVLEAGSPGSIYNNCSVPLTAARDGAVAILQELLDHGAENAVKA 212
302 LPVWASNTASCSPGLYLAAYVGHLCDFRLLLLHGHADPDYNCDDQGLLARVPRPTLLEI 361
302 LPVWASNTASCSPGLYLAAYVGHLCDFRLLLLHGHADPDYNCDDQGLLARVPRPTLLEI 361
213 LPVWASNTASCSPGLYLAAYVGHLCDFRLLLLHGHADPDYNCDDQGLLARVPRPTLLEI 272
213 LPVWASNTASCSPGLYLAAYVGHLCDFRLLLLHGHADPDYNCDDQGLLARVPRPTLLEI 272
362 LHNCEPEYIQLLIDFGANIYLPSSLDTSDDDKGIALLQARATPRSLLSQVRLVVR 421
362 LHNCEPEYIQLLIDFGANIYLPSSLDTSDDDKGIALLQARATPRSLLSQVRLVVR 421

Db 273 LHHNCEPEYIQLIDFGANIYPSLSLSDTSQDDKGIALLOARATPRSLLSQVRLVWR 332

QY 422 ALCAQGPQAINOLDIPPMILSYLKHQ 449
|||||

Db 333 ALCAQGPQAINOLDIPPMILSYLKHQ 360
|||||

RESULT 4
US-09-908-193-12
; Sequence 12, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHIMKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
; APPLICANT: PADIGARU, MURALIDHARA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 21402-062
; CURRENT APPLICATION NUMBER: US/09/908,193
; PRIOR FILING DATE: 2001-07-18
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: 60/220,273
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/221,650
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/221,233
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/220,912
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/218,875
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,870
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,901
; PRIOR FILING DATE: 2000-07-18
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: NOVX polypeptide
; OTHER INFORMATION: with homology to ankryrin repeat-containing-
; OTHER INFORMATION: (ASB-1) like protein
US-09-908-193-12

Query Match 69.5%; Score 1629; DB 10; Length 351;
Best Local Similarity 97.5%; Pred. No. 1.8e-159;
Matches 318; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 132 MRVLQAKNLMIDITKIFSLLOPDKEEDTDTEEKQALNQAVYDNDSTLDQLLRERY 191
|||||

Db 26 MRVLQAKNLMIDITKIFSLLOPDKEEDTDTEEKQALNQAVYDNDSTLDQLLRERY 85
|||||

QY 192 KRFNRSRGVPGTPLRLAASVGHLSCLQVLLAHGADVSLDVKAQTPFTAVSHGLD 251
|||||

Db 86 KRFNRSRGVPGTPLRLAASVGHLSCLQVLLAHGADVSLDVKAQTPFTAVSHGLD 145
|||||

QY 252 CVRVLLAAGAPSGSIYNNCSPLVTAARDGAVAILQELLDHGAENAVKAKLPWASNIAS 311
|||||

Db 146 CVRVLLAAGAPSGSIYNNCSPLVTAARDGAVAILQELLDHGAENAVKAKLPWASNIAS 205
|||||

QY 312 CSGPLVLAAYGHLDLFRLLHLLHGADPDYNTDQGLLARVPRPTLLEICLHHNCEPEYI 371
|||||

Db 206 CSGPLVLAAYGHLDLFRLLHLLHGADPDYNTDQGLLARVPRPTLLEICLHHNCEPEYI 265
|||||

QY 372 QLLIDFGANIYPSLSLSDTSQDDKGIALLOAR-----ATPRSLLSQVRLVWRAL 423
|||||

Db 266 QLLIDFGANIYPSLSLSDTSQDDKGIALLOARGLFLLAVATPRSLLSQVRLVWRAL 325
|||||

QY 424 COAQGPQAINOLDIPPMILSYLKHQ 449
|||||

Db 326 COAQGPQAINOLDIPPMILSYLKHQ 351

RESULT 5
US-09-908-193-36
; Sequence 36, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHIMKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
; APPLICANT: PADIGARU, MURALIDHARA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 21402-062
; CURRENT APPLICATION NUMBER: US/09/908,193
; PRIOR FILING DATE: 2001-07-18
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: 60/220,273
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/221,650
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/221,233
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/220,912
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/218,875
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,870
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,901
; PRIOR FILING DATE: 2000-07-18
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 36
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-908-193-36

Query Match 61.2%; Score 1435; DB 10; Length 308;
Best Local Similarity 88.6%; Pred. No. 1.7e-139;
Matches 273; Conservative 17; Mismatches 18; Indels 0; Gaps 0;

QY 141 NMLMDITKIFSLLOPDKEEDTDTEEKQALNQAVYDNDSTLDQLLRERYKRFNRSRG 200
|||||

Db 1 NMLMDITKIFSLLOPDKEEDTDTEEKQALNQAVYDNDSTLDQLLRERYKRFNRSRG 60
|||||

QY 201 WGVPGTPLRLAASVGHLSCLQVLLAHGADVSLDVKAQTPFTAVSHGLDCVRVLEAG 260
|||||

Db 61 WGVPGTPLRLAASVGHLSCLQVLLAHGADVSLDVKAQTPFTAVSHGLDCVRVLEAG 120
|||||

QY 261 ASPGSGSIYNNCSPLVTAARDGAVAILQELLDHGAENAVKAKLPWASNIASCSGPLYLA 320
|||||

Db 121 ACPGSGSIYNNCSPLVTAARDGAVAILQELLDHGAENAVKAKLPWASNIASCSGPLYLA 180
|||||

QY 321 VYGHLDLFRLLHLLHGADPDYNTDQGLLARVPRPTLLEICLHHNCEPEYIQLLDIFGAN 380
|||||

Db 181 VYGHLDLFRLLHLLHGADPDYNTDQGLLARVPRPTLLEICLHHNCEPEYIQLLDIFGAN 240
|||||

QY 381 IYLPISLSDTSQDDKGIALLOARATPRSLLSQVRLVWRALCOAGOPQAINOLDIPPM 440
|||||

Db 241 IYLPISLSDTSQDDKGIALLOARATPRSLLSQVRLVWRALCOAGOPQAINOLDIPPM 300
|||||

QY 441 LISYLKHQ 448
|||||

Db 301 LISYLKHQ 308
|||||

RESULT 6
US-09-908-193-37
; Sequence 37, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 1, 2003, 07:39:07 ; Search time 4237 Seconds
(without alignments)
4335.242 Million cell updates/sec

Title: US-09-941-831A-20
Perfect score: 2343
Sequence: 1 MRESGWKLIDPISDFGRMG.....QAINQLDIPPMISYKQL 449

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xl
-Q=/cgn2_1/USPTO_spool_p/US09941831/runat_01122003_073527_7829/app_query.fasta_1.647
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=200 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09941831@cgn_1_1_3745@runat_01122003_073527_7829 -NCPU=6 -ICPU=3
-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*

- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1690	72.1	1222	6	AX418418 Sequence
2	1681	71.7	1269	9	AK096896 Homo sapi
3	1629	69.5	1069	6	AX418416 Sequence
c	1606.5	68.6	104534	9	AL356317 Human DNA
	1599	68.2	927	9	AF403030 Homo sapi
6	1459	62.3	1262	10	AF403040 Mus muscu
c	1391	59.4	191093	10	AL671765
	1376.5	58.7	194079	2	AC109692 Rattus no
9	1376.5	58.7	225982	2	AC108644
10	899.5	38.4	240816	5	AL844570 Zebrafish
11	675.5	28.8	2653	9	BC012399 Homo sapi
12	642.5	27.4	1345	6	AX406008 Sequence
13	499	21.3	1008	9	AF156777 Homo sapi
14	499	21.3	1340	9	BC014528 Homo sapi
15	481	20.5	1008	10	AF155352 Mus muscu
16	481	20.5	5892	10	AK122456 Mus muscu
17	480	20.5	2019	6	AR262497 Sequence
18	480	20.5	2019	6	AX022304 Sequence
19	480	20.5	2019	6	AX030771 Sequence
20	480	20.5	2019	6	BD007458 Remedies
21	454	19.4	6518	9	AB032972 Homo sapi
22	434.5	18.5	2354	5	BC053411 Danio rer
23	431.5	18.4	1312	10	BC049699 Mus muscu
24	424.5	18.1	2353	6	AX086038 Sequence
25	421	18.0	3065	10	BC020019 Mus muscu
26	409	17.5	2337	9	AK093237 Homo sapi
27	409	17.5	4069	9	BC040012 Homo sapi
28	408	17.4	1866	9	AF406619 Homo sapi
29	400	17.1	2754	5	BC044359 Danio rer
30	397.5	17.0	2817	10	BC032254 Mus muscu
31	395	16.9	4131	9	HSB803324
32	394	16.8	2482	6	AX399878 Sequence
33	378.5	16.2	3059	3	BT003304 Drosophil
34	378.5	16.2	3209	3	AF181629 Drosophil
c	335	14.3	77274	2	AC005672 Drosophil
	335	14.3	154895	2	AC020329 Drosophil
c	335	14.3	181955	3	AC008304 Drosophil
c	335	14.3	218565	3	AC009257 Drosophil
c	335	14.3	299331	3	AE003460 Drosophil
	313.5	13.4	42724	3	CEP53A2
40	313.5	13.0	1214	3	AK113728
41	305.5	12.7	1966	6	BD156744 Primer fo
42	297	12.7	1966	6	AK001816 Homo sapi
43	297	12.7	2397	9	BC052990 Homo sapi
44	296.5	12.7	2397	9	BC040432 Homo sapi
45	296.5	12.7	3158	9	BC040432

ALIGNMENTS

RESULT 1

```

AX418418
LOCUS AX418418 1222 bp DNA linear PAT 18-JUN-2002
DEFINITION Sequence 13 from Patent WO0206329.
ACCESSION AX418418
VERSION AX418418.1 GI:21523342
KEYWORDS unidentified
SOURCE unidentified
ORGANISM unclassified.
REFERENCE
1 Rastelli, L., Shimkets, R.A., Zerkusen, B., Malyankar, U.M. and Padigaru, M.
HUMAN polynucleotides and polypeptides encoded thereby
PATENT: WO 0206329-A 13 24-JAN-2002;
CUREGEN CORPORATION (US)
LOCATION/Qualifiers
1. 1222
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/note="NOVX nucleic acid with homology to ankyrin repeat-containing- (ASB-1) like protein"
BASE COUNT 276 a 352 c 278 g 316 t
ORIGIN
Alignment Scores:
Pred. No.: 4,57e-147 Length: 1222
Score: 1690.00 Matches: 326
Percent Similarity: 99.70% Conservative: 1
Best Local Similarity: 99.39% Mismatches: 1
Query Match: 72.13% Indels: 0
DB: 6 Gaps: 0
US-09-941-831A-20 (1-449) x AX418418 (1-1222)
QY 122 ProLysLeuThrHisPheLeuValLeuLeuMetArgIleValLeuGlnLeuAlaLysMet 141
DB 102 CCACAGATCGCATTTCTTGTGCTCATATGAGATAGTCTCCATTAATGCCAAGATG 161
QY 142 AsnLeuMetAspIleThrLysLeuPheSerLeuLeuGlnProAspLysGluGluAsp 161
DB 162 AACCTCATGGACATCACCAAGATCTTCTCCCTCGACGCCGACAAAGGAGGAGGAC 221
QY 162 ThrAspThrGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThr 181
DB 222 ACTGACACAGAGAGAGAGAGGCTCTCAATCAAGCAGTGTATGACACGACTCCTATATC 281
QY 182 LeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGlyTyr 201
DB 282 TTGGACCACTTTTGGCCAGGAGCGTTTCAACAGTTTTCATCACAGCAGGAGTGCTGG 341
QY 202 GlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGln 221
DB 342 GGTGTTCTCGGACACCTTGGCTTGGCTGCTTCTTATGGCCACTTGAGCTGTTGCAA 401
QY 222 ValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeu 241
DB 402 GTCTCTTAGCCATCGTGTATGTGACAGCTTGGATGTCAAGGCACAGACGCCACTT 461
QY 242 PheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAla 261
DB 462 TTCACGTCTGTACGTATGCCATCTGGACTGTGTACGTGTGCTTTTGAAGCTGGTGCC 521
QY 262 SerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGly 281
DB 522 TCTCTGGTGGTAGCATCTACAACAACCTGTCTCCGCTGCACAGCTGCCGTGATAGT 581
QY 282 AlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnValLysAlaLys 301
DB 582 GCTGTTGCTATCTCTGACGAGGCTCTTAGACCATGTGGTGCAGGCCAACGCTCAAGCTAA 641
QY 302 LeuProValTyrAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaVal 321

```

```

642 CTACCAGTCTGGCATCAAAACATAGCTTCTATGTTCTGGCCCTCTATTTGGCCGAGTC 701
322 TyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsn 341
702 TACGGGCACCTGGACTGTTTCGCCTGCTTTTGTCTCCACGGGGCAGACCTGACTACAC 761
342 CysThrAspGlnGlyLeuLeuAlaArgValProAspProArgThrLeuLeuGluLys 361
762 TGCACTGACCGGGCTATTTGGCTCGTGTCCCAAGACCCCGCACCTCTCTTGAATCTGC 821
362 LeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuLeuAspPheGlyAlaAsnIle 381
822 CTCCATCATTAATTGTGAGCCAGAGTATATCCAGCTGTTAATCGATTGTTGGTCTAATATC 881
382 TyrLeuProSerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeuLeu 401
882 TACCTTCCATCTCTCTCCCTTGACCTGACCTCACAGATGATAAAGGCATTCATTTGCTG 941
402 LeuGlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgAsp 421
942 CTACAGGCCGAGCCACTCCACGGTCACTTCTATCACAGTCCGTTAGTCTCCGACA 1001
422 AlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProMetLeu 441
1002 GCCTTGTGCGAGCTGGCCAGCCACCAAGCATCAACAGCTGGATATCTCCCATGTTG 1061
442 IleSerTyrLeuLysHisGlnLeu 449
1062 ATTAGCTACCTAAAACACCAACTG 1085
RESULT 2
AK096896 1289 bp mRNA linear PRI 15-JUL-2002
LOCUS to Homo sapiens cDNA FLJ39577 fis, clone SKMUS2002920, weakly similar
DEFINITION to Homo sapiens ASB-1 protein mRNA.
ACCESSION AK096896
VERSION AK096896.1 GI:21756490
KEYWORDS oligo capping, fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Tashiro, H., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuno, Y., Nagai, K. and Isogai, F.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1269)
AUTHORS Isogai, T. and Yamamoto, J.
JOURNAL Direct Submission
COMMENT Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.
FEATURES
source
1. 1269
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="SKMUS2002920"

```

```

/tissue_type="skeletal muscle"
/clone_lib="SKMUS2"
/notes="cloning vector: pME18SFL3"
174. .1130
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC04888.1"
/db_xref="GI:21756491"
/translation="MRIVLQAKNKLMDITKIFSLQPKDEEDTDTTEKQALNQAVY
DNDSYTLQQLLRQRYKFIINSRSGWGVPPTPLRAASYGHLSCLOVLLAHGADVDSL
DKVATPLFTAVSHGLDVSCLLEAGAPGSGSYNNCSPLVTAARDGVAIIIEQLD
HGAEANVAKKLPVANSNTSCSPGLYAAVYGHLDCEFLILLHCAQPDYNTDQGLLA
RVPRRTLEICLHNCEPEYIQLIDFGANYILPSSLTSDODDKGIALLLQARAT
FRSLSQVRLVRRALCOGQPDQAINLDIPPMILISYKHQL"
BASE COUNT 288 a 370 c 305 g 306 t
ORIGIN

Alignment Scores:
Pred. No.: 3,29e-146 Length: 1269
Score: 1681.00 Matches: 325
Percent Similarity: 99.69% Conservativeness: 0
Best Local Similarity: 99.69% Mismatches: 1
Query Match: 71.75% Indels: 0
DB: 9 Gaps: 0

US-09-941-831A-20 (1-449) x AK096896 (1-1269)

QY 124 IleTrpHisPheLeuValLeuIleMetArgIleValLeuGlnLeuAlaLysMetAsnLeu 143
DB 150 ATCTGGCATTCCTTGCTCTATATAGATAGTGTCTCCANTTAGCCAGATGAACCTC 209

QY 144 MetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGluAspThrAsp 163
DB 210 ATGGACATCAACAGATCTCTCCCTCTGAGCCGACAGAGGAGGAGGAGGAGGAGGAG 269

QY 164 ThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThrLeuAsp 183
DB 270 ACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 329

QY 184 GlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGlyTyrGlyVal 203
DB 330 CAGCTTTTGGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 389

QY 204 ProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeu 223
DB 390 CTTGGGACACCCCTTGGCTTGGCTTCTTATGGCCACTTGAGCTGTTTGCAGTCTCTC 449

QY 224 LeuAlaHisGlyAlaAspValAspSerLeuLeuAspValLysAlaGlnThrProLeuPheThr 243
DB 450 TTAGCCCATGGTGTGATGTTGACAGCTTGGATGTCAAGGCACAGAGCCCACTTTTCACT 509

QY 244 AlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerPro 263
DB 510 GCTGTCAGTCATGGCCCATCTGACTGTGTAGTGTGCTTTTGGAGCTGGTGCCCTCTCT 569

QY 264 GlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaVal 283
DB 570 GGTGTAGCATCTACAACACTGTTCTCCCGTCTCACAGCTGCCCGTGTGTGTGTGTGT 629

QY 284 AlaIleLeuGlnGluLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeuPro 303
DB 630 GCTATCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 689

QY 304 ValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAlaValTyrGly 323
DB 690 GTCTGGGCATCAACATAGCTTCATGTCTTGGCCCCCTCTATTTGGCCGAGCTCTACGG 749

QY 324 HisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsnCysThr 343
DB 750 CACCTGGACTGTTTCGGCTGCTTTTGTCTCACGGGAGGAGGAGGAGGAGGAGGAGGAGG 809

QY 344 AspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIleCysLeuHis 363

```



```
Db 10836 TTCACTGCTGTACGTATCGCCATCTGGAGCTGTGTGTGCTTTTGGAAAGCTGTGTGCC 10777
Qy 262 SerProGlyGlySerIleTyrAenAenCysSerProValLeuThrAlaAlaAraGAspGly 281
Db 10776 TCTCTGTTGGTGGTAGCATCTACAACTGTTCTCCGCTGCTACAGCTGCCGTGTATGTT 10717
Qy 282 AlaValAlaIleLeuGlnLeuLeuAraGAspHisGlyAlaGluAlaAraValAlaLys 301
Db 10716 GCTGTTGCTATCTGCAGGAGCTCTTAGACCATGTTGCAGAGCCCAACGTCRAAGCTAAA 10657
Qy 302 LeuProValTrrAlaSerAenIleAlaSerCysSerGlyProLeuTyrLeuAlaAlaVal 321
Db 10656 CTACCAAGCTCTGGGCATCAACATAGCTTATGTTCTGGCCCCCTCTATTGTCGCGAGTC 10597
Qy 322 TyrGlyHisLeuAraGAspHisGlyAlaGluAlaAraValAlaLys 341
Db 10596 TAGCGGACCTGAGCTGTTTCCGCCCTGCTTTTCTCCAGGGGACACCTGACTACAC 10537
Qy 342 CysThrAraGlnGlyLeuLeuAlaAraGValProAraGProAraGThrLeuLeuGluLeuCys 361
Db 10536 TGCACTGACCAAGGCTATTTGGCTGCTGTCCTCCAGAGCCCGCACCTCTCTTGAATCTGC 10477
Qy 362 LeuHisHisAraGlnGlyProGluTyrIleGlnLeuLeuAraGAspPheGlyAlaAraVal 381
Db 10476 CTCATCATATAATTTGTGAGCCAGATATATCCAGCTGTTAATCATTTTGTGTCTAATATC 10417
Qy 382 TyrLeuProSerLeuSerLeuAraGAspLeuThrSerGlnAraGAspGlyIleAlaLeuLeu 401
Db 10416 TACCTTCCATCTCTCTCCCTTGACCTGCTTCAAGATGATTAAGCATTTGATTTGCTG 10357
Qy 402 LeuGlnAlaAraGla 406
Db 10356 CTACAGGCCCCGAGG-TGAGTGGTGAACAGGCGACGCTCATAGGCCACCTGCTTGGAGT 10298
Qy 406 ----- 406
Db 10297 CCCAGGGCTCTAAATCAGCTATTAAGCAATCTCTGAGGTGCACTGGAGAAAACCTCCATC 10238
Qy 406 ----- 406
Db 10237 TTCCTCCCTCTCTGGGGCCCCAGTAGAGAGGGGTTATGATTTCCATGTTAGAGAAA 10178
Qy 406 ----- 406
Db 10177 GGTCACTAAGATAAGTGTGGTGGGGTGGGAGCATTAGCCCAATTTCTGGTCTTGCT 10118
Qy 406 ----- 406
Db 10117 CTGCCAATGACTGTTTATTTCTTTGAAAGGATAGAGCAATTAGGGCTGTAGTCCAGTGG 10058
Qy 406 ----- 406
Db 10057 AGAGTGGATGAGGTTGTGGGAGAAATTTTCTGATTAATGCTCTCATGACCTGTTT 9998
Qy 407 -----ThrProAraGserLeuLeuSerGlnValAraGLeuValValAraGArg 421
Db 9997 CTTCTTGTCTAGCCACTCCAGGCTCATTCTATCAGAGTCCGTTTGTGCTGCTCCAG 9938
Qy 422 AlaLeuCysGlnAlaGlyGlnProGlnAlaIleAraGlnLeuAraGAspLeuProMetLeu 441
Db 9937 GCCTTGTGCGAGCTGGCCAGCCACAGCCATCAACAGCTGATATTTCTCCCATGTTG 9878
Qy 442 IleSerTyrLeuLeuHisGlnLeu 449
Db 9877 ATTAGCTACTAAACACCAACTG 9854
RESULT 5
AF403030 927 bp mRNA linear PRI 02-JAN-2002
LOCUS Homo sapiens ankyrin repeat domain-containing SOCS box protein
DEFINITION Abb-12 mRNA, complete cds.
ACCESSION AF403030
VERSION AF403030.1 GI:18034079
```

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-941-831A-20 (1-449) x AF403030 (1-927)

Qy 141 MetAsnLeuMetAspIleThrIlePheSerLeuLeuGlnProAspIleGluGlu 160

Db 1 ATGAACCTCATGGACATCACCAAGATCTTCTCCCTCTGCGAGCCGACAGGAGGAG 60

Qy 161 AspThrAspThrGluGluLysGlnAlaLeuAraGlnAlaValTyrAspAraAspSerTyr 180

Db 61 GACACTGACACAGAGAGAGAGGAGGCTCTCAATCAAGCAGTGTATGACCAACGACTCTAT 120

Qy 181 ThrLeuAspGlnLeuLeuAraGlnGluArgTyrLysArgPheIleAsnSerArgSerGly 200

Db 121 ACTTTGGACCACTTTTGGCCAGGAGCTTACAAACGTTTCATCAACAGCAGGAGTGGC 180

Qy 201 TrpGlyValProGlyThrProLeuArgLeuAlaAraSerTyrGlyHisLeuSerCysLeu 220

Db 181 TGGGGTGTTCCTGGGACACCTTGGCTTGGCTTCTTATGGCCACTTGAGCTCTTTG 240

Qy 221 GlnValLeuLeuAlaHisGlyAlaAraGAspValAspSerLeuAspValLysAlaGlnThrPro 240

Db 241 CAAGTCTCTTAGCCCATGCTGCTGATTTGACAGCTTGGATGTCAAGGACACAGAGCCCA 300

Qy 241 LeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuAlaGly 260

Db 301 CTTTTCACCTGCTGCTCAGTCTATGGCCATCTGGAGTGTGTACGTGTGCTTTTGAAGTGT 360

Qy 261 AlaSerProGlyGlySerIleTyrAraGlnCysSerProValLeuThrAlaAlaAraGAsp 280

Db 361 GCCTCTCTCTGGTGGTAGCATCTACAACTGTTCTCCCGTGTCTACAGCTGCCCCGTGAT 420

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1. (bases 1 to 927)

Kile, B.T. and Nicola, N.A.

SOCS box proteins

Unpublished

2. (bases 1 to 927)

Kile, B.T. and Nicola, N.A.

Direct Submission

Submitted (27-JUL-2001) Cancer & Hematology, The Walter & Eliza

Hall Institute of Medical Research, Royal Melbourne Hospital,

Parkville, Victoria 3050, Australia

Location/Qualifiers

1. 927

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

1. 927

/codon_start=1

/product="ankyrin repeat domain-containing SOCS box

protein Abb-12"

/protein_id="AA157349.1"

/db_xref="GI:18034080"

/translation="NNLMDITKIFSLLOPDKEEDTTEKQALNAQAVYDNDSTLDDQ

LLRORYKRFINSRSGWVPGTPLRAASYGHLSCLOVLLAHGADVSLDVKAOPTPLF

TAVSHGLDCTVRVLEAGASPGSGSYNNCSFVLTAAAGDAVAILQELLDHGAENYKA

KLFWASASNYSCSGPLVLAAYVGHDCFRLLLLHAGADPDYNTDGLLARVRPRPTLL

ETCLHNGEPEYIQLIDFGANIYPLSLDLTSDQDKGIALLQARATPRSLLSQVR

LVRRALCAQAGQPOAINQIDIPPMILSYLKHQ"

203 a 274 c 216 g 234 t

8.56e-139 Length: 927

1599.00 Matches: 308

100.00% Conservative: 0

100.00% Mismatches: 0

68.25% Indels: 0

9 Gaps: 0

QY 281 GlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnValLysAla 300
 Db 421 GGTGCTGTTGCTATCTCGCAGGAGCTCTTAGACCATGGTGCAGAGCCAAAGCT 480
 QY 301 LysLeuProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAla 320
 Db 481 AAACCTACCAAGTCTGGGCATCAACATAGCTTCATGTTCTGGCCCCCTCTATTGGCCGCA 540
 QY 321 ValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyr 340
 Db 541 GTCTAGGGCAGCTGAGTCTTCCGCTGCTTTTCTCCAGGGGAGACCCCTGACTAC 600
 QY 341 AsnCysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIle 360
 Db 601 AACTGCACTGACCAAGGCTATTGGCTCGTGTCCCAAGACCCCGCACCCTCTTGAATC 660
 QY 361 CysLeuHisHisLeuAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsn 380
 Db 661 TGCCTCCATCAATTTGTGAGCAGAGTATATCCAGCTGTTATCGATTGTTGGTCTAAT 720
 QY 381 IleTyrLeuProSerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeu 400
 Db 721 ATCTACTTCCATCTCTCTCCCTTGACCTGACCTCAAGATGATAAAGCATTTGCATTG 780
 QY 401 LeuLeuGlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValAlaArg 420
 Db 781 CTGCTACAGGCCCGAGCCACTCCACGGTCACTTCTATCACAGTCCGTTAGTCTCGC 840
 QY 421 ArgAlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProMet 440
 Db 841 AGAGCTTGTGCGAGGCTGCCAGCCACCAAGCCATCAACAGCTGGATATCTCCCATG 900
 QY 441 LeuIleSerTyrLeuLysHisGln 448
 Db 901 TTGATTAGCTACCTAAACACCAA 924

RESULT 6

AF403040 1262 bp mRNA linear ROD 02-JAN-2002
 LOCUS Mus musculus ankyrin repeat domain-containing SOCS box protein
 DEFINITION
 AF403040 Asb-12 mRNA, complete cds.
 ACCESSION
 VERSION AF403040.1 GI:18034099
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Kille, B.T. and Nicola, N.A.
 SOCS box proteins
 TITLE Unpublished
 JOURNAL Kille, B.T. and Nicola, N.A.
 REFERENCE 2 (bases 1 to 1262)
 AUTHORS Kille, B.T. and Nicola, N.A.
 TITLE Direct Submission
 JOURNAL Submitted (27-JUL-2001) Cancer & Hematology, The Walter & Eliza
 Hall Institute of Medical Research, Royal Melbourne Hospital,
 Parkville, Victoria 3050, Australia
 Location/Qualifiers
 FEATURES
 source 1..1262
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 198..1124
 /codon_start=1
 /product="ankyrin repeat domain-containing SOCS box
 protein Asb-12"
 /protein_id="AA157359.1"
 /db_xref="GI:18034100"
 translations="NM1MDIAKIFSLQPEKEEDTGTGCKQALNOAVYDNDPSCTLDH
 LHQERYKRFINSRSGWIGPITRLAAGYHNCVKVLEHGVADVDSLDVRAQTPLF
 TAVSHGLECVRMLEAGACPSGIYNGSPVLTASRDGAFALQLLGHGAENYKA
 KLPVWASNTASCSPGLYLAAYVGHLCDFRLLLYLGADPDYNCCTDQGLLRVPQPTLL

CDS

198..1124
 /product="ankyrin repeat domain-containing SOCS box
 protein Asb-12"
 /protein_id="AA157359.1"
 /db_xref="GI:18034100"
 translations="NM1MDIAKIFSLQPEKEEDTGTGCKQALNOAVYDNDPSCTLDH
 LHQERYKRFINSRSGWIGPITRLAAGYHNCVKVLEHGVADVDSLDVRAQTPLF
 TAVSHGLECVRMLEAGACPSGIYNGSPVLTASRDGAFALQLLGHGAENYKA
 KLPVWASNTASCSPGLYLAAYVGHLCDFRLLLYLGADPDYNCCTDQGLLRVPQPTLL

BASE COUNT 318 a 338 c 272 g 334 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.24e-125 Length: 1262
 Score: 1459.00 Matches: 282
 Percent Similarity: 92.94% Conservative: 21
 Best Local Similarity: 85.50% Mismatches: 22
 Query Match: 62.27% Indels: 2
 DB: 10 Gaps: 1
 US-09-941-831A-20 (1-449) x AF403040 (1-1262)
 QY 124 IleTrpHisPheLeuValLeuIleMetArgIleValLeu--GlnLeuAlaLysMetAsn 142
 Db 145 ATCTGGAATTTCTCTCTCTGCAATGAGATA-GTGTATTTCGAATAAACACCAAGATGAAC 203
 QY 143 LeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGluAspThr 162
 Db 204 CTCATGGATATCGCCAGATCTTCTCTCTGCAACCTGAAAGGAGGAGGAGACAT 263
 QY 163 AspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThrLeu 182
 Db 264 GGCACCGGGGAAAAACAGGCTCTCAATCAAGCTGTATATGACATCTCTGTACCCCTG 323
 QY 183 AspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGlyTrpGly 202
 Db 324 GACCACTTCTACACCGAGAACGTTATAACGGTTTCATCAACAGCAGGAGTGGGGT 383
 QY 203 ValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnVal 222
 Db 384 ATACCTGGAACACCTTGGCTTGGCAGCTTCTTATGGTCACTTAAATTTGTGAGGTC 443
 QY 223 LeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPhe 242
 Db 444 CTCCTGGAACATGGTGTGATGTTGATAGTTGATGTCAGGAGCTTAATGTCAGAACTAAT 503
 QY 243 ThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSer 262
 Db 504 ACTGTGTGAGCCACGGTCTCATCTGGAGTGGTGAGAAATGCTTTTAGAAGCTGGTCTGT 563
 QY 263 ProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAla 282
 Db 564 CTTAGTGGTAGCATCTACAAACAATGCTCTCTGTCTCACCTGCTCTACGATGGGGCT 623
 QY 283 ValAlaIleLeuGlnGluLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeu 302
 Db 624 TTTGGCATCTTACAGGAGCTCTAGGGCATGTCGCGAGGCTTAATGTCAGAACTAAT 683
 QY 303 ProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaValTyr 322
 Db 684 CCAGTCTGGGCTCAAAATATAGCTTCTGCTGCCCCCTCTATCTGCTGCGAGTCTAT 743
 QY 323 GlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsnCys 342
 Db 744 GGGCACTTGTATGTTTTCGCTGCTTGTCTATGGGGGAGATCCTCATTAACAATGC 803
 QY 343 ThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIleCysLeu 362
 Db 804 ACTGACCGAGGCTTTTAAGTGTGTTCACAGCTCGCACACTCTCTTGAATCTGCTT 863
 QY 363 HisHisAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsnIleTyr 382
 Db 864 CATCAATTTGTGAGCAGAGTACATCCAGCTTTTAATAGATTTTGGAGCTAACATCTAC 923
 QY 383 LeuProSerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeuLeuLeu 402
 Db 924 CTTCCATCTCTCTCTGTGGACCAACTTCCAAAGATGATAAAGGCAATCAAAATGCTGCTA 983
 QY 403 GlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgAla 422

Db 984 CAAGCCCGAGCACTCCAGGTCACCTCTGTGTCCAGACCCGTTAGTTATTCGCCGAGATCC 1043

Qy 423 LeuCysGlnAlaglyGlnProGlnAlaIleAsnGlnLeuAspIleProPrometLeuIle 442
 Db 1044 CTCTGCCGGGCGAACCAAGTCACAAAGCCACCGACAGCTGGATATCCCCCTGTGTGATT 1103

Qy 443 SerTyrLeuIlysHisGln 448
 Db 1104 AGCTACCTCAACATCAA 1121

RESULT 7
 AL671765/c
 LOCUS
 DEFINITION Mouse DNA sequence from clone RP23-100L23 on chromosome X, complete sequence.
 ACCESSION AL671765
 VERSION AL671765.6 GI:21666377
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 191093)
 Lloyd,D.
 Direct Submission
 Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Jul 2, 2002 this sequence version replaced gi:21614613.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones. Where differences are found these are compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-100L23 is from the RP23-23 Mouse PAC library constructed by the group of Pieter de Jong.
 For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBAC3.6.

FEATURES
 source Location/Qualifiers
 1..191093
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone="RP23-100L23"
 /clone_lib="RPCi-23"
 BASE COUNT 58421 a 38555 c 38626 g 55491 t
 ORIGIN

Alignment Scores:
 Pred. No.: 2,34e-116 Length: 191093
 Score: 1391.00 Matches: 283
 Percent Similarity: 66.30% Conservative: 22
 Best Local Similarity: 61.52% Mismatches: 22

```

QY 405 ----- 405
Db 90162 ACTAAGATGAGGTATGTAGAAAGCATTTATACATCTCTGGTCTTAGATCTCTCTAATGAC 90103
QY 405 ----- 405
Db 90102 GGACCCCTTTGATGCCCTGATAGGCTCAGTATATGTTATCATGCAATTGGGAAGATGGTT 90043
QY 405 ----- 405
Db 90042 GAGACAGTCTTAGAAGGAGGAGGCGCATTACTATATAGCCATGCCAGATAGATGTACA 89983
QY 406 -----AlaThrPro 408
Db 89982 CTTGGAATTACACTGAGAATTGATGCTCATAATCTGTTTCTCATTTGTCGACGCCACTCCA 89923
QY 409 ArgSerLeuLeuSerGlnValArgLeuValValArgAlaLeuLeuCysGlnAlaGlyGln 428
Db 89922 CGGTCACTCTCTGCCAGACCCGTTAGTATCCGAGATCCCTCTGCCGGCGCAACACAG 89863
QY 429 ProGlnAlaLeuAsnGlnLeuAspIleProMetLeuIleSerTyrLeuLysHisGln 448
Db 89862 TCACAGGCCACCGACCGAGCTGGATATCCCTGTTGTTAGTACTCAACATCAA 89803

RESULT 8
AC109692 194079 bp DNA linear HTG 11-OCT-2002
LOCUS Rattus norvegicus clone CH230-268J15, *** SEQUENCING IN PROGRESS
DEFINITION ***
ACCESSION AC109692
VERSION HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 194079)
AUTHORS Muzny,D,Marie, Metzker,M, Lee, Abranzon,S, Adams,C, Alder,J,
Allen,C, Allen,H, Alsbrooks,S, Amin,A, Anguiano,D,
Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,
Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F,
Biswalo,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M,
Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E,
Cardenas,V, Carter,K, Cavazos,I, Caesar,H, Center,A,
Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,
Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,
Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D,
Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K,
Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K,
Egan,A, Escoto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G,
Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,
Frazer,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M,
Greggeorgis,E, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W,
Gunaratne,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K,
Harvey,Y, Haylak,P, Hawes,A, Henderson,N, Hernandez,J,
Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M,
Hollins,B, Howells,S, Hulyk,S, Hume,J, Idlebird,D, Jackson,A,
Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A,
Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C,
Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J,
Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J,
Lorenshewa,L, Louised,H, Lozado,R,J, Lu,X, Ma,J,
Maheshwari,M, Mahindarne,M, Mahmoud,M, Malloy,K, Mangum,A,
Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E,
Mawhney,S, McLeod,M,P, McNeill,T,Z, Meenen,E,
Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S,
Morgan,M, Morris,K, Morris,S, Munidas,M, Murphy,M, Nair,L,
Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Parks,K,
Nwaoketeme,O, Okwuonu,G, Olarnpunsagoon,A, Pal,S, Parks,K,
Pasternak,S, Paul,H, Perez,A, Perez,L, Pfannkuch,C,
Plopper,F, Poindexter,A, Popovic,D, Primus,E, Pu,L-L,
Puazo,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R,

```

```

TITLE
JOURNAL
AUTHORS
REFERENCE
JOURNAL
COMMENT
Rat Genome Sequencing Consortium.
Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza Houston, TX 77030, USA
3 (bases 1 to 194079)
Rat Genome Sequencing Consortium.
Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza Houston, TX 77030, USA
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
Center: Genome Center
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GQBH
Center clone name: CH230-268J15
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 167200 bases at least Q40
Consensus quality: 170449 bases at least Q30
Consensus quality: 172661 bases at least Q20
Estimated insert size: 186192; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 194079: contig of 194079 bp in length.
Location/Qualifiers
1. 194079
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
FEATURES
source

```

Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneddie,M., Sodergren,E., Song,X.-Z., Sorelle,R., Soosa,J.,
Steinle,M., Strong,N., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczek,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 194079)
Worley,K.C.

Direct Submission
Submitted (07-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza Houston, TX 77030, USA

3 (bases 1 to 194079)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza Houston, TX 77030, USA

On Oct 9, 2002 this sequence version replaced gi:21737980.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

Center: Genome Center
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GQBH
Center clone name: CH230-268J15

----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 167200 bases at least Q40
Consensus quality: 170449 bases at least Q30
Consensus quality: 172661 bases at least Q20

Estimated insert size: 186192; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 194079: contig of 194079 bp in length.

Location/Qualifiers
1. 194079
/organism="Rattus norvegicus"
/mol_type="genomic DNA"

FEATURES
source

```

/db_xref="taxon:10116"
/clone="CH230-268J15"
1..1257
/note="wgs_end_extension
clone_end:T7"
misc_feature
6621..7521
/note="clone_boundary
clone_end:T7"
site:Mbol
end sequence:RXAE056TJ"
110101..112444
/note="wgs_contig"
192195..193062
/note="clone_boundary
clone_end:Sp6
site:Mbol
end sequence:RXAE056TV"
BASE COUNT 48623 a 34416 c 35611 g 55333 t 20096 others
ORIGIN

Alignment Scores:
Pred. No.: 5.28e-115 Length: 194079
Score: 1376.50 Matches: 283
Percent Similarity: 62.76% Conservative: 22
Best Local Similarity: 58.23% Mismatches: 25
Query Match: 58.75% Indels: 156
DB: 2 Gaps: 3

US-09-941-831A-20 (1-449) x AC109692 (1-194079)
QY 119 AspThrArgProLysIleTyrHis----- 126
Db 152178 GAGAGTAAACATAGACTGTGGCAATGTAATGATGCTTAGTCTTACACCTCCCTTTT 152237
QY 127 -----PheLeuValLeuIleMetArgIleValLeu---GlnLeu 138
Db 152238 CATTACTCTCAGATCGGGAATTTCTCTCTTTTACAAATGAGATAGCTCTATTCCAATAA 152297
QY 139 AlaLysMetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGlu 158
Db 152298 ACCAGATGAACTCATGATATCGCAAGATCTTCTCTTCTGCAACGAGAAAAGAG 152357
QY 159 GluGluAspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAsp 178
Db 152358 GAGGAGGACGCCGACACAGCGAAAAGCAGGCTCTCAATCAAGCTGTATATGACATGAC 152417
QY 179 SerTyrThrLeuAspGlnLeuLeuArgGlnGluArgTyrIlysrghPheIleAsnSerArg 198
Db 152418 TCTTGTAACCTTGACACCACTTCTACACCGAACGTTATAAACGGTTTCATCAACAGCAGG 152477
QY 199 SerGlyTyrGlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSer 218
Db 152478 AGTGGCTGGGGTATACCTGGAACACCTTTCGGCTTGGCAGCTCTTATGGTCACTTAGAT 152537
QY 219 CysLeuGlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGln 238
Db 152538 TGTGTCAAAAGTCTCTCTGGAACATGCTGCTGATGTTGATAGCTTGGATGTCAAGACACA 152597
QY 239 ThrProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGlu 258
Db 152598 ACACCACTTTTCACTGCTGTAGTACGCTCATCTGGAGTGTGTGAGAATCTTTTAGAA 152657
QY 259 AlaGlyAlaSerProGlyLysSerIleTyrAsnAsnCysSerProValLeuThrAlaAla 278
Db 152658 GCTGTGCTGCTGCTAGTGTAGTACATCAACAATGTCTCTCTCTCTCTCTCTCTCTCTCT 152717
QY 279 ArgAspGlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnVal 298
Db 152718 CGTGATGGGGCCCTTTGCCATCTTACAGGAGCTCTTAGGGCATGGTGCAGAGCTAATGTC 152777
QY 299 LysAlaLysLeuProValThrAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeu 318
Db 152778 AAGGCGAAATCTGCAGTCTGGCATCAACATAGCTTCTATGTTCTGGCCCCCTCTATTG 152837

/description="Rattus norvegicus clone CH230-59AG, *** SEQUENCING IN PROGRESS ***"
AC108644
LOCUS
DEFINITION
Rattus norvegicus clone CH230-59AG, *** SEQUENCING IN PROGRESS ***
8 unordered pieces.
AC108644
VERSION
AC108644.5 GI:24818595
HTGS PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 225982)
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,

```

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, J., Cleaveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gunaratne, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaragob, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Huiy, S., Hume, J., Idelbird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowals, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Loresunehwa, L., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munitasa, M., Murphy, M., Naif, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okunolu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pu, M., Pu, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zimm, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 225982)
Worley, K.C.

Direct Submission
Submitted (31-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 225982)

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (09-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 9, 2002 this sequence version replaced gi:23267762.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature

table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPSJ
Center clone name: CH230-59A6
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 211822 bases at least Q40
Consensus quality: 215372 bases at least Q30
Consensus quality: 217787 bases at least Q20
Estimated insert size: 216316; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 32687: contig of 32687 bp in length
* 32688 32787: gap of unknown length
* 32788 102606: contig of 69819 bp in length
* 102607 102706: gap of unknown length
* 102707 123303: contig of 20597 bp in length
* 123304 123403: gap of unknown length
* 123404 123409: contig of 88306 bp in length
* 123410 212409: gap of unknown length
* 212410 213673: contig of 1264 bp in length
* 213674 213773: gap of unknown length
* 213774 214949: contig of 1176 bp in length
* 214950 215049: gap of unknown length
* 215050 217109: contig of 2060 bp in length
* 217110 217209: gap of unknown length
* 217210 225982: contig of 8773 bp in length.
----- Location/Qualifiers
1. 225982
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-59A6"
complement(4374..5073)
/note="clone_boundary"
clone_end:17
site:ECORI
end sequence:BH331578"
32788..35656
/note="wgs_contig"
81431..82325
/note="clone_boundary"
clone_end:Sp6
site:ECORI
end sequence:BH331579"

BASE COUNT 65353 a 44744 c 44903 g 64320 t 6662 others
ORIGIN
Alignment Scores:
Pred. No.: 6,49e-115 Length: 225982
Score: 1376.50 Matches: 283
Percent Similarity: 62.76% Conservative: 22
Best Local Similarity: 58.23% Mismatches: 25
Query Match: 58.75% Indels: 156
DB: 2 Gaps: 3
US-09-941-831A-20 (1-449) x AC108644 (1-225982)

Qy	119	AspThrArgProLysIleTrpHis-----	126
Db	129983	GAGAGTAAACATAGAGTCTGGCATTTGATCAATAATGATCGTTAGCTTTCACACCCTCCCTTTTTT	130042
Qy	127	-----PheLeuValLeuIleMetArgIleValLeu---GlnLeu	138
Db	130043	CATTACTCTCAGATCGGGAATTTCTCTCTTTTACAATGAGAATAGCTCTATTCCAATAA	130102
Qy	139	AlaLysMetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGlu	158
Db	130103	ACCAGAATGAACCTCATGGATATGCCAAGATCTTCTCTCTCTGCACCAACCAAAAAGAG	130162
Qy	159	GluGluAspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyzAspAsnAsp	178
Db	130163	GAGGAGGACGCCGACACAGGCGAAAAAGAGGCTCTCAATCAAGCTGTATATGACAATGAC	130222
Qy	179	SerTyrrThrLeuAspGlnLeuLeuArgGlnGluArgTyrrLysArgPheIleAsnSerArg	198
Db	130223	TCTTGTAACCTTGGACCACTTCTACACGAGAACGTTATAAACGGTTTCATCAACAGCAGG	130282
Qy	199	SerGlyTrpGlyValProGlyThrProLeuArgLeuAlaAlaSerTyrrGlyHisLeuSer	218
Db	130283	AGTGGCTGGGGTATACCTGGAAACACCTTGGCTTGGCAGCCTCTTATGGTCACCTTAGAT	130342
Qy	219	CysLeuGlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGln	238
Db	130343	TGTGTGAAGTCTCTCTGGAACATGCTGATGTTGATAGCTTGGATGTCAAAGCACAA	130402
Qy	239	ThrProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGlu	258
Db	130403	ACACCACATTTTCACTGCTGTAGTCACCGCTCATCTGGAGTGTGTGAGAATACTTTTAGAA	130462
Qy	259	AlaGlyAlaSerProGlyGlySerIleTyrrAsnAsnCysSerProValLeuThrAlaAla	278
Db	130463	GCTGGTGGCTGCTCTAGTGGTAGCATCTACAACAATTGCTCTCTCTCTCACTGCCCTCA	130522
Qy	279	ArgAspGlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnVal	298
Db	130523	CGTGATGGGGCTTTTGGCCATCTTACAGGAGCTCTTAGGCGATGGTGCAGAGGCTAATGTC	130582
Qy	299	LysAlaLysLeuProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrrLeu	318
Db	130583	AAGCGAAACTGGCAGCTCTGGGCATCAACACATAGCTTCATGTTCTTGGCCCCCTCTATTG	130642
Qy	319	AlaAlaValTyrrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspPro	338
Db	130643	GCTGCAGTCTAAGTCATCTTGATTGTTCCGCCTGCTTTTGCTCTATGGGCGCATCTCT	130702
Qy	339	AspTyrrAsnCysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeu	358
Db	130703	GATTACAACCTGATTGACAGGCCCTCTTAAGTCGAGTTCACACAGCCTCGCACCTCTT	130762
Qy	359	GluIleCysLeuHisHisAsnCysGluProGluTyrrIleGlnLeuLeuIleAspPheGly	378
Db	130763	GAATATCGCCTTCATCATAAATTGTGAGCGAGAGTACATACAGCTTTTAATAGATTTTGA	130822
Qy	379	AlaAsnIleTyrrLeuProSerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyLe	398
Db	130823	CGGAACATCTACCTCCCACATCTTACCTTTGGACCCAAACATCTCAAGATGATAAAGGCATC	130882
Qy	399	AlaLeuLeuLeuGlnAlaArg-----	405
Db	130883	AAGTTGCTGCTACAAGCCCGAGGTGAATAGTATCAGGCATACCTCATAGGCTTCTTTAT	130942
Qy	405	-----	405
Db	130943	TGGAGGTCTCAGGGTTTCCAAATCATAGACCTTCCCACTGGGTTGACCTAGAAATGGCTC	131002
Qy	405	-----	405
Db	131003	CATCATCTCTGTGGTCCCTGTGGAGAAAAAGTTAAAGATTTTCAAGAAGTGAAAGGT	131062
Qy	405	-----	405

Db	131063	TACTAAGATGAGGTGTGTAGAAAGCATTTTATACAGTCTCTGGCCCTTAGATCTGCGCAATGA	131122
Qy	405	-----	405
Db	131123	CTCACTCTTTTGATGCCCTGAGCTTCTCTTCTATAGGCCTTGGTGATGTTTACCATGCAT	131182
Qy	405	-----	405
Db	131183	TTGGAGATGATTGTGACAATACTTTAGACGGGAGAAGGCATTATATAGCCATGGCAGAAT	131242
Qy	405	-----	405
Db	131243	AGATAGACACTTGGACAGAAATAAGTACACAGAGAAATCGATCCTCATATTTGTTTCTCAT	131302
Qy	406	-----AlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgAlaLe	423
Db	131303	TTTGGCAGCTACTCCACGGTCACTCTCTGTCCAGGCCCGCTTAGTTATCCGACAGATCCCT	131362
Qy	423	uCyseGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProMetLeuIleSe	443
Db	131363	GTGCCGAGCAACACGAGCCACAGCATCGACCATGCGATATCCCCCTGTGTGATTAG	131422
Qy	443	rTyrLeuLysHisGln	448
Db	131423	CTACCTCAAGCATCAA	131438
RESULT	10		
LOCUS	AL844570	240816 bp	DNA linear VRT 11-FEB-2003
DEFINITION	Zebrafish DNA sequence from clone DKEY-20D18, complete sequence.		
ACCESSION	AL844570		
VERSION	AL844570.9	GI:28300586	
KEYWORDS	HTG.		
SOURCE	Danio rerio (zebrafish)		
ORGANISM	Danio rerio		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio. 1 (bases 1 to 240816)		
AUTHORS	Sehra, H.		
TITLE	Direct Submission		
JOURNAL	Submitted (11-FEB-2003), Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfsh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Feb 10, 2003 this sequence version replaced gi:27652710.		
COMMENT	----- Genome Center Center: Wellcome Trust Sanger Institute Web site: http://www.sanger.ac.uk Contact: zfsh-help@sanger.ac.uk		

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/c_elegans/wormpep Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhongqiang Bao and Sean Eddy, submitted), and those beginning 'dr'.

were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.Projects/D_rerio/fishmask.shtml
 DKEY-20D18 is from a Zebrafish BAC library
 VECTOR: pindisoBAC-5

FEATURES

```

source
  1. 240816
  /organism="Danio rerio"
  /mol_type="genomic DNA"
  /db_xref="taxon:7955"
  /clone="DKEY-20D18"
  /clone_lib="DanioKey"

BASE COUNT 76749 a 43368 c 43554 g 77145 t
ORIGIN

Alignment Scores:
Pred. No.: 1,17e-70 Length: 240816
Score: 899.50 Matches: 178
Percent Similarity: 75.08% Conservative: 45
Best Local Similarity: 59.93% Mismatches: 61
Query Match: 38.39% Indels: 13
DB: 5 Gaps: 4

US-09-941-831A-20 (1-449) x AL844570 (1-240816)

QY 141 MetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGlu 160
DB 18737 ATGAGTTTCATGGACATCTTAAGATCTTTCTCTTCAACCC---AAAGAGGATGAG 18793
QY 161 AspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyAspAsnAspSerTy 180
DB 18794 GAGGAAGACACCAACTCGAGTCAGCACTAAATCAAGCTGTCTTCACAGACGACGACAA 18853
QY 181 ThrLeuAspGlnLeuLeuArgGlnGluAArgTyLysArgPheIleAsnSerArgSerGly 200
DB 18854 CTCCTTACTGAAGTGTGTCTCAGGAACAATATAGAAAGTGCATCAACACCGCAGTGT 18913
QY 201 TrpGlyValProGlyThrProLeuArgLeuAlaAlaSerTyGlyHisLeuSerCysLeu 220
DB 18914 TGGGGATCCCATCTCACTCCGACGCGCGCTGCTCAGGGTCACCTACGCTGTCTG 18973
QY 221 GlnValLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrPro 240
DB 18974 GAGGTCCTCTGCTCAGCGGCGAGAGTGCAGACGCTGGATGCTCAAGAGCTCAACACCA 19033
QY 241 LeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGly 260
DB 19034 CTTTTCACAGCCGCTCTGCGCAACACATAGATTGTGTGCTGCTTTATTAAGAGCTGC 19093
QY 261 AlaSerProGlyGlySerIleTyAsnAsnCysSerProValLeuThrAlaAlaArgAsp 280
DB 19094 GCAGACCTTANTGGAAGCCACATACACTGCTCCCTGTGTGACTGCGCCAGAGAG 19153
QY 281 GlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnValLysAla 300
DB 19154 GGGGATGTGGACATCTCTGAAGAGAGCTCTTCAATATATGGCGCCGATGAGAGCTCAAGCC 19213
QY 301 LysLeuProValTrpAlaSerAniIleAlaSerCysSerGlyProLeuTyIleuAlaAla 320
DB 19214 AAAATGCCAGACTGGGCGCTTAAATGCCCACAGCTTGGCGAGGACCACTGTACATTTCCAG 19273
QY 321 ValTyIleGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTy 340
DB 19274 GTGTATGTCACTAGTGTGTTTCAATTTGTTGCTGAGTTTGGAGCCCAACCCAGACTAC 19333
QY 341 AsnCysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIle 360
DB 19334 AACTGCACGTGAGGAAAAGATGCTGGCCAGGATCAAGAGCCATAAACTGTTTGGAGATG 19393
QY 361 CysLeuHisIleAsnCysGluProGluTyIleGlnLeuLeuIleAspPheGlyAlaAsn 380
DB 19394 TGTCTGAGACATGGCTGTGGGCTGGGAATACATTCAGCTGCTCATAGACTTTGGAGCAAT 19453
QY 381 IleTyIleProSerLeuSerLeuAspLeuThrSerGlnAspLysGlyIleAlaLeu 400

```

```

Db 19454 GTTATCTGCCCTTTAGTTGGAGATAAAACCCAGTGAAGTGAACGAGTAGTCTT 19513
QY 401 LeuLeuGlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArg 420
DB 19514 CTGCTTAAAGAGAGAGGTGAG-----TCTCAATTACAC----- 19546
QY 421 ArgAlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIle 437
DB 19547 -----TTGTGTCATCTTTTAAG---ACCGGTTTCAACATTTTAGACATT 19588

```

RESULT 11

```

BC012399
LOCUS BC012399
DEFINITION Homo sapiens, Similar to hypothetical protein FLJ20126, clone
MGC:8876 IMAGE:3864128, mRNA, complete cds.
ACCESSION BC012399
VERSION BC012399.1 GI:15214550
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Strausberg,R.
Direct Submission
Submitted (15-AUG-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalona@bcm.tmc.edu.
Villalón, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 21 Row: j Column: 23
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7020020.

FEATURES

```

source
  1. 2653
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="MGC:8876 IMAGE:3864128"
  /tissue_type="Ovary, adenocarcinoma"
  /clone_lib="NIH_MGC_66"
  /lab_host="DH10B"
  /note="Vector: pCMV-SPORT6"
  69. 2183
  /codon_start=1
  /product="similar to hypothetical protein FLJ20126"
  /protein_id="AAH12399.1"
  /translation="MDHITVPKVENVKLVDRYVSKKPKANGILYLTATHLIYVEASGAA
  RKETALHHTATVEKLPITSLGCLPTLRCKNFRVAHFLDSDLVCHVEYIILKLSQ
  PALPDLAFSNPKSSKEMESGWLIDPSDFGRMGI PNRNWTITDANRNVICST
  YPPELVKPSVTLGTGVGSSKRSKRPVLYLYKNNAAICRSQPSLGSPTRCVD
  DELLLEAISQTNPGSQFVVDTRPKLNANRAAGKYENEDNVTANIRFRFVGIENI
  HVMRSLOKLLEVCLEKLPPTSEFSLGSLGRHKAIMDAGIFITKAVVEKASV
  LVHCSGWDRTAQVCVASILDDPFRTPFKGLMILIEKWI SMGHKFSQRCCHLDGDS
  KEVSPIFTQFLDCIQWLMEQPPCAFEFNFLIEHDHVFVSCQFGNLCQCKREDLI

```

CDS

RYEKTHSVNPLVORPKDFRNPFLYKGYTMVGLNPFSTVPIYNIQFWMYRPFKGLQ
PKQSMLESLLEIKKORAMLETDVHELEKLVKRDPEPERICTCSOLGNLSHLGSLP
TNPLGPMGLINDLNTLMENGLTSLRREGGLRAQMDQVKSQCADLHNNCEIVGSLRINI
SGDVGISEAMGISDGMCTEATGFSKDLGICGMDISEATGISGNLISEARGFSGDM
GILGDTGISKASTKEADYSKHQ

BASE COUNT 749 a 566 c 645 g 693 t
ORIGIN

Alignment Scores:
Pred. No.: 1,46e-52 Length: 2653
Score: 675.50 Matches: 189
Percent Similarity: 50.95% Conservative: 52
Best Local Similarity: 39.96% Mismatches: 126
Query Match: 28.83% Indels: 107
DB: 9 Gaps: 19

US-09-941-831A-20 (1-449) x BC012399 (1-2653)

QY 1 MetArgGluSerGlyTrpLysLeuIleAspProIleSerAspPheGlyArgMetGlyIle 20
DB 432 ATGAGGGAAGTGGATGGAAAGTGAATGACCAATATCAGACTTTGGGCGTATGGGAATA 491
QY 21 ProAsnArgAsnTrpThrIleThrAspAlaAsnArgAsnTyrGluIleCysSerThrTyr 40
DB 492 CCCAACAGAACTGGACCATACAGATGCCAACAGAACTATCAGATATGCAGACCTAC 551
QY 41 ProProGluIleValValProLysSerValThrLeuGlyThrValValGlySerSerLys 60
DB 552 CTTCTCGAATAGTGTCTTAATCTGTTACCTTGGGAACGGTGTGGAAAGTTCANAG 611
QY 61 PheArgSerLysGluArgValProValLeuSerTyrLeuTyrLysGluAsnAlaAla 80
DB 612 TTCAGAAAGTAAGACGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 671
QY 81 IleCysArgCysSerGlnProLeuSerGlyPheTyrThrArgCysValAspAspGluLeu 100
DB 672 ATTTCGCGCTAGCCAGCGCTCTCTCTGGAATTTTACACTCGCTGTGTAGATGATGAGCTC 731
QY 101 LeuLeuGluAlaIleSerGlnThrAsnProGlySerGlnPheMetTyrValValAspThr 120
DB 732 TTGTTGGAGCCATTAGCCAAACACCCAGGAGCCAGTTATGATGATGTTAGACACA 791
QY 121 ArgProLysIle----- 124
DB 792 AGACCAAGTGAATGCCATGGCCACCGCAGCAGCTGGGAGGGGTATGAAATGAAGAC 851
QY 125 -----TrpHisPheLeu-----ValLeuIleMetArgIle 134
DB 852 AACTATGCCAACATTCGCTTCAGATTTCATGGGCATTGAGAACATCCATGTAATGCGGAGC 911
QY 135 ValLeuGlnLeuAlaLysMetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGln 154
DB 912 AGTCTCGAG-----AACTCTTGGAAAGTTGTGAATGAAACT 950
QY 155 ProAspLysGluGluAspThrAspThrGluGluLysGlnAlaLeuAsn-----Gln 172
DB 951 CCACAACTAGTGAATTTCTAGCGCCCTGGAGAGCTCAGGGTGGTTAGACACATTAA 1010
QY 173 AlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeuLeuArgGlnGluArgTyrLys 192
DB 1011 GCTATTATGATGCTGGAATTTTC---ATTACAAAGGCAGTGAAGGTAGAAAGGCCAGT 1067
QY 193 ArgPheIleAsnSerArgSerGlyTrpGlyValProGlyThrProLeuArgLeuAlaAla 212
DB 1068 GTCTTAGTCCATGTTCTGATGATGGGACCGCAGCAGCAAGCTGTCTCAGTGGCTAGC 1127
QY 213 -----SerTyrGlyHisLeuSerCysLeuGlnValLeuLeu----- 224
DB 1128 ATCTCTCTAGATCCATTTTATAGGACATTCAGAGGACTCATGCTTGTATAGAGGAA 1187
QY 225 -----AlaHisGlyAlaAspValAspSer 232
DB 1188 TGGATATCATCGGCCACAAAGTTTTTCCCAAGAGGTGTGGCCAC---CTCGATGGGACTCT 1244

QY 233 LeuAspValLysAlaGlnThrProLeuPheThrAlaValSerHisGlyHisLeuAspCys 252
DB 1245 AAGAAGTG-----TCCCTATCTTCACCCAG-----TTCCTAGACTGT 1283
QY 253 ValArgValLeuLeuGlu-AlaGlyAlaSerProGlyGlySerIleTyrAsnAsnCys-- 271
DB 1284 ATCTGGCAATTAAATGGAACAGTTTCCTGTGCTTTGAGTTTAAATGAAATTCCTCTGCTG 1343
QY 272 -----SerPro-----ValLeuThrAlaAla 279
DB 1344 GAGATTTCATGACCATGTTTCTCTCTGCCAGTTTGGAAATTCCTTGTGTAATGCCAGAAG 1403
QY 279 GAspGlyAlaVal-----AlaIleLeuGlnGluLeu----- 290
DB 1404 GATCGGGAAGATCTAAGAGTCTATGAGAAAACACATTTCTGTGTGGCTTCTTGTGGTTCAG 1463
QY 291 -----AspHisGlyAlaGluAlaAsnValLysAlaLysLeuProValTrpAlaSerAs 308
DB 1464 AGAAACACAGACTTCAGGAACCTCTCTATAAGGCTTCACCTATGATGGGTACTCAAT 1523
QY 308 nIleAlaSerCysSerGlyProLeuTyrLeuAlaValTyrGlyHisLeuAspCysph 328
DB 1524 CCTAGTACTGTGC-----CCTACAACTTCAGTTCTGTGTGGATGTATAACCGCT-- 1575
QY 328 eArgLeuLeuLeuHisGlyAlaAspProAspTyrAsnCysThrAspGlnGlyLeuLe 348
DB 1576 -----TTGACAAAGGGCTGCAGCCCAAGCAGATATGCTA-GAGAGC-----CT 1618
QY 348 uAlaArgValProArgProArgThrLeuLeuGluIleCysLeuHis----- 363
DB 1619 CTGGAAATTAAAGAACAGAGAGCAATGCTGGAGACAGATGTCATGAACATAGAAAAGAA 1678
QY 364 -----HisAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAla 380
DB 1679 ACTAAAGTCCGTGATGAGCCACCAAGAGATCTGTACCTCTCTCAATTAGGA---AA 1735
QY 380 nIleTyrLeuProSerLeuSerLeuAspLeuSer 392
DB 1736 CATATTATCCAGCATCTGGGAAGTCTTTGACCAAT 1772
RESULT 12
LOCUS AX406008 1345 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 423 from Patent WO0222660.
ACCESSION AX406008
VERSION AX406008.1 GI:21439455
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Tang, Y.T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q.A., Ren, F.,
Xue, A.J., Yang, Y., Wehrman, T. and Drmanac, R.T.
TITLE Novel nucleic acids and polypeptides
JOURNAL Patent: WO 0222660-A 423 21-MAR-2002;
HYSEQ, INC. (US)
FEATURES
Location/Qualifiers
1..1345
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
92..1204
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD35115.1"
/db_xref="GI:21439456"
/translation="MDHITVPKVENVKLVDRYVSKKPPANGILYLTLTATHLIYVEASGAA
RRTWIAHHIATVEKPIITSGCPLTUCRKNFRVAHFVLDLVCHEVYISLLKSQ
PALPDLFYAFSNFKSKEMRESGKLLDIPISDFRGMGIPNRNITITDANRYEICST
YPRIVVPSKVTGLTVGSSKRSKRPVLSLYLKENNAICRLSPLSEPTRCVD
DELLLEISQTPGSGQPMYVVDTRPKAVKVEKASVLVHCSGDGWDRTAQVCSVASILLD

```
PPYRTFKGLMILIEKEMSGHFKSQRCGLHDGDSKVSPIPTQFLDCIWLMEQFPC
APEFNENFLLETHDHFVSCQFCNFCNCKQKREDLR"
BASE COUNT      380 a  300 c  321 g  344 t
ORIGIN
Alignment Scores:
Pred. No.:      6.65e-50      Length:      1345
Score:          642.50      Matches:      145
Percent Similarity: 61.34%      Conservative: 20
Best Local Similarity: 53.90%      Mismatches: 33
Query Match:      27.42%      Indels:      71
DB:              6          Gaps:      8

US-09-941-831A-20 (1-449) x AX406008 (1-1345)
QY 1 MetArgGluSerGlyTrpLysLeuLeuAspProLysSerAspPheClyArgMetGlyIle 20
Db 455 ATGAGGGAAGTGGATGGGAACCTGATTGACCAATATCAGATTTGGGGGTATGGGAATA 514
QY 21 ProAsnArgAsnTrpThrIleThrAspAlaAsnArgAsnTyrGluIleCysSerThrTyr 40
Db 515 CCCACAGAACTGGACCATACAGATGCCACAGAACTATGAGATATGACGACCTAC 574
QY 41 ProProGluLeuValValProLysSerValThrLeuGlyThrValValGlySerSerLys 60
Db 575 CCTCTGAAATAGTGGTCTCTAAATCTGTTACTTGGGAACGGTGGTGGAAAGTTCAAAG 634
QY 61 PheArgSerLysGluArgValProValLeuSerTyrLeuTyrLysGluAsnAsnAlaAla 80
Db 635 TTCAGAAAGTAAAGAACGTGCTGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 694
QY 81 IleCysArgCysSerGlnProLeuSerGlyPheTyrThrArgCysValAspAspGluLeu 100
Db 695 ATTGGCGGTTGAGCCAGCCTCTCTCTGAAATTTACCTCGCTGTGTAGATGATGAGCTC 754
QY 101 LeuLeuGluAlaIleSerGlnThrAsnProGlySerGlnPheMetTyrValValAspThr 120
Db 755 TTGTTGGAGCCATTAGCCAAACAAACCCAGGAGCCAGTTTATGTATGTTGTAGACACA 814
QY 121 ArgProLysIleTrpHisPheLeuValLeuIleMetArgIleValLeuGlnLeuAlaLys 140
Db 815 AGACCAAAG-----GCAGTGAAGTAGAAGAAAG 841
QY 141 MetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGluGlu 160
Db 842 GCCAGTGTCTTA----- 853
QY 161 AspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyr 180
Db 854 -----GTCCATTGTTCTGATGGATGG 874
QY 181 -----ThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArg 193
Db 875 GACCGCACACACAGCTCTGCTCAGTGGGTAGCATCTCTCTAGATCCATTTTATAGGACA 934
QY 194 Phe-----IleAsnSerArgSerGlyTyrGlyValProGlyThrProLeuArg 209
Db 935 TTCAAAGGACTCATGATCTTGATAGAGAGGAATGGATATCCATGGGCCAC-----AAG 988
QY 210 LeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAlaHisGlyAlaAsp 229
Db 989 TTTTCCCAAAGGTGGCCACCTC-----GAT 1015
QY 230 ValAspSerLeuAspValLysAlaGlnThrProLeuPheThrAlaValSerHisGlyHis 249
Db 1016 GGGGACTCTAAAGAGTG-----TCCCTATCTTCACCCAG-----TTC 1054
QY 250 LeuAspCysValArgValLeuLeuGlu 258
Db 1055 CTAGACTGTATCTGGCAATTAATGGAA 1081
RESULT 13
AF156777
```

```
LOCUS      AF156777              1008 bp      mRNA      linear      PRI 12-DEC-2000
DEFINITION Homo sapiens ASB-1 protein mRNA, complete cds.
ACCESSION   AF156777
VERSION     AF156777.1 GI:5306061
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1008)
AUTHORS     Kile,B.T., Viney,E.M., Willson,T.A., Brodnicki,T.C., Cancilla,M.R.,
Herlihy,A.S., Croker,B.A., Baca,M., Nicola,N.A., Hilton,D.J. and
Alexander,W.S.
TITLE       Cloning and characterization of the genes encoding the ankyrin
            repeat and SOCS box-containing proteins Asb-1, Asb-2, Asb-3 and
            Asb-4
JOURNAL     Gene 258 (1-2), 31-41 (2000)
MEDLINE     20564172
PUBMED      11111040
REFERENCE   2 (bases 1 to 1008)
AUTHORS     Kile,B.T.
TITLE       Direct Submission
JOURNAL     Submitted (07-JUN-1999) Cancer & Hematology, The Walter & Eliza
            Hall Institute of Medical Research, Royal Parade, Parkville, VIC
            3052, Australia
FEATURES
            source
            1..1008
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="zt10f03; IMAGE:712733"
            CDS
            1..1008
                /note="ankyrin repeat-containing protein; contains SOCS
                box 1"
                /codon_start=1
                /product="ASB-1 protein"
                /protein_id="AAD41894.1"
                /db_xref="GI:5306062"
                /translation="MAEGSPDGRAGPGSAGRNLEKWLREQFCDFHLECHEDTRLHDA
                AYVGLDQLTSLQEEYSRINSEKSVNCCGWLPCTPLRIATAGHGSQVDFLRKGA
                EVDLVKGGTALYVAVVNGHLESTQILLEAGDENGSRHHRSTPTVHRSVRGRADIL
                KALTRYGADVNVNHLTPDQVPRFSRLTSLVCPLYISAAVHNLCQFRLLLAGANP
                DFCNPGVPTQGYRGSPGVMDAIVLRHGCFAAFVSLVFGANLNLVKWESLGPESR
                GRKVDPEALQVFKEARSPVPTLLCLCRVAVRVALGKRLHLIPSLPLPDPIKKFLH
                E"
BASE COUNT      201 a  293 c  310 g  204 t
ORIGIN
Alignment Scores:
Pred. No.:      9.01e-37      Length:      1008
Score:          499.00      Matches:      119
Percent Similarity: 57.8%      Conservative: 57
Best Local Similarity: 39.14%      Mismatches: 102
Query Match:      21.30%      Indels:      26
DB:              9          Gaps:      9

US-09-941-831A-20 (1-449) x AF156777 (1-1008)
QY 166 GluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeu 185
Db 109 GAGGACACAGAGCTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 168
QY 186 LeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGlyTyr-----Gly 202
Db 169 TTGCAAGAGGAGAGCTACCGGAGCCGACATCAACAGAGAGTCTGTCTGTGTGTGTGTGTGT 228
QY 203 ValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnVal 222
Db 229 CTCCTCTGCACACCGTTGCGAATGTCGGCCATCGAGCCATCGAGGAGCTGTGTGGACTTC 288
QY 223 LeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPhe 242
Db 289 CTCATCCGGAAGGGGGCCGAGGTGGATCTGGTGGACGTAAGAGGACAGACGCGCCTGTAT 348
```



```

Db      838 AGAAGATGATCCGAGGCTTGCAGGTCTTTAAAGAGGCCAGAGTATTCCAGGACC 897
Qy      411 LeuLeuSerGlnValArgLeuValValArgAlaLeuCysGlnAlaGlyGlnProGln 430
Db      898 TTGCTGAGTTGTGCCGGGTGGCTGTGAGAAGAGCTCTTGGCAAATAC---CGACTGCAT 954
Qy      431 AlaIleAsnGlnLeuAspIleProProMetLeuIleSerTyrLeuLysHisGln 448
Db      955 CTGGTTCCCTCGCTGCCGCTGCCAGACCCCAAAAGNAGTTTTTGTATTATGAG 1008

```

Search completed: December 1, 2003, 09:17:07
 Job time : 4616 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 1, 2003, 07:38:12 ; Search time 353 Seconds
(without alignments)
3433.563 Million cell updates/sec

Title: US-09-941-831A-20

Perfect score: 2343

Sequence: 1 MRESGWKLDPISDFGRMGI.....QAINQLDIPPMISYLKHLQ 449

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer_p2n.model -DEV=xl
-O=/cgn2_1/USPro.spool_p/US09941831/runat_01122003_073526_7819/app.query.fasta_1.647
-DB=N Geneseq 19Jun03 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09941831@cgn_1_1_396@runat_01122003_073526_7819 -NCPUS=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq_19Jun03:*

1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2343	100.0	1878	22	AAH46857	Human serine/threose
2	2343	100.0	1878	22	AAH78730	Human HIBCJ89 seri
3	2343	100.0	1878	22	AAS33372	DNA encoding human
c	2331	99.5	2530	22	AAS33264	DNA encoding human
5	1690	72.1	1222	24	ABK11107	DNA encoding human
6	1680	71.7	1127	24	ABZ11533	DNA encoding human
7	1629	69.5	1069	24	ABK11106	DNA encoding human
c	672	28.7	8214	23	AAS77257	Novel human coding
9	642.5	27.4	1345	24	ABN60012	DNA encoding human
c	499	21.3	1266	22	AAS33071	Human prostate exp
11	499	21.3	1433	23	ABV23231	Human prostate exp
12	499	21.3	1433	23	ABV29074	Human cytoskeleton
13	499	21.3	1509	25	AAD49617	DNA encoding human
14	499	21.3	1540	22	AAS33289	Mus musculus SOCS7
15	480	20.5	2019	19	AAV38672	Human phosphatase
16	424.5	18.1	2353	22	AAF63581	Human cDNA SEQ ID
17	394	16.8	2482	24	ABQ93336	Drosophila melanog
18	378.5	16.2	2786	23	ABL06295	Drosophila melanog
19	378.5	16.2	3047	23	ABL20003	Drosophila melanog
20	378.5	16.2	3169	23	ABL17297	Drosophila melanog
c	366	15.6	408	23	ABV61004	Human prostate exp
c	335	14.3	8397	23	ABL20002	Drosophila melanog
c	335	14.3	9648	23	ABL06294	Drosophila melanog
c	335	14.3	9648	23	ABL17296	Drosophila melanog
25	297	12.7	1966	22	AAH14752	Human cDNA sequenc
26	293.5	12.5	545	23	AAS73111	DNA encoding novel
27	287	12.2	3382	24	ABK35507	Human cDNA differe
28	281	12.0	2710	22	AAK52009	Human polynucleoti
29	281	12.0	4330	24	AAD40741	Human kinase and p
30	275.5	11.8	2575	23	AAS73882	DNA encoding novel
31	273	11.7	2346	22	AAK52993	Human polynucleoti
32	270.5	11.5	1622	21	AAK36830	DNA encoding a su
c	264.5	11.3	2764	22	AAH17868	Human cDNA sequenc
34	264	11.3	3411	18	AAT91998	Human tyrosine pho
35	261.5	11.2	1206	22	AAI58158	Human polynucleoti
36	261.5	11.2	3398	24	AAI71054	Benign prostatic h
37	257.5	11.0	1024	24	ABQ54306	Human ovarian anti
38	252.5	10.8	1869	22	AAH13777	Human cDNA sequenc
39	252.5	10.8	2263	21	AAZ36833	DNA encoding a su
40	246	10.5	802	22	AAH07276	Human cDNA clone (
41	246	10.5	1538	24	AAH93131	Human C/SKARP-1 pr
42	244.5	10.4	401	21	AAA77749	DNA encoding huma
43	244.5	10.4	401	22	AAI28487	Colon tumour relat
44	244.5	10.4	401	25	ABZ32673	Human colon tumour
45	241	10.3	1565	22	ABA09163	Human cytokine sig

ALIGNMENTS

RESULT 1
AAH46857

ID AAH46857 standard; cDNA; 1878 BP.

XX AAH46857;

XX DT 25-SEP-2001 (first entry)

XX Human serine/threonine phosphatase encoding cDNA (clone ID HIBCJ89).

XX Serine/threonine phosphatase; nootropic; neuroprotective; cytostatic;
KW dermatological; immunosuppressive; antiinflammatory; antibacterial; ss;
KW anti-HIV; antiparkinsonian; antischistosomal; antianemic; antiarthritic;
KW antirheumatic; virucide; hepatotropic; cerebroprotective; vulnary;
KW antiinflammatory; nephrotropic; gene therapy; vaccine.
XX Homo sapiens.
XX OS

QY 401 LeuLeuGlnAlaAArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArg 420
 |||||
 Db 1567 CTGCTACAGCCGCGAGCACTCACGGTCACCTCTATCACAGTCGGTTAGCTCGTCGCG 1626
 |||||
 QY 421 ArgAlaLeuCysGlnAlaGlyGlnProGlnAlaAlaAsnGlnLeuAspLeuProMet 440
 |||||
 Db 1627 AGAGCGTTGTGCGAGCTGCGCCAGGCACCAAGCCATCAACAGCTGGATATCTCCCATG 1686
 |||||
 QY 441 LeuLeuSerTyrLeuLeuHisGlnLeu 449
 |||||
 Db 1687 TTGATTAGCTACCTAAACACCAACTG 1713
 |||||

RESULT 2

AAH78730

ID AAH78730 standard; cDNA; 1878 BP.

XX AC AAH78730;

XX DT 03-DEC-2001 (first entry)

XX DE Human HIBCJ89 serine/threonine phosphatase cDNA sequence.

XX KW Human; HIBCJ89; ss; serine/threonine phosphatase; PSPase; vaccine;
 gene therapy; PSPase expression; PSPase expression; PSPase modulation;
 KW immune disorder; autoimmune disorder; Wiscott-Aldrich syndrome;
 KW Chediak-Higashi syndrome; Hashimoto's thyroiditis; multiple sclerosis;
 KW inflammation; Crohn's disease; inflammatory bowel disease; appendicitis;
 KW rheumatoid arthritis; cellular proliferative disorder; lymphoma;
 KW lung cancer; intestinal cancer; cardiovascular disorder; aneurysm;
 KW Scimitar syndrome; Ebstein's anomaly.
 XX KW Homo sapiens.

FH Key Location/Qualifiers

FT CDS 367..1716

FT /*tag= a

FT /product= "HIBCJ89 PSPase protein"

XX WO200164703-A1.

XX PD 07-SEP-2001.

XX PF 28-FEB-2001; 2001WO-US06256.

XX PR 02-MAR-2000; 2000US-0186350.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ebner R, Ruben SM;

XX DR WPI; 2001-530113/58.

XX DR P-PSDB; AAG77803.

XX Nucleic acids encoding serine/threonine phosphatase polypeptides,
 PT useful for preventing, diagnosing and/or treating, e.g. Crohn's
 PT disease, lung cancer and Scimitar syndrome -

XX PS Claim 1; Page 308; 335pp; English.

XX CC The present sequence represents the specifically claimed human cDNA clone
 CC HIBCJ89 which shares homology with members of the serine/threonine family
 CC of phosphatases. The invention comprises novel human serine/threonine
 CC phosphatase (PSPase) polypeptides and polynucleotides which may be used
 CC in the prevention (vaccine), diagnosis and treatment (gene therapy) of
 CC diseases associated with inappropriate PSPase expression. The PSPase
 CC polynucleotides of the invention may be used as DNA probes to detect and
 CC quantitate the presence of similar nucleic acids in samples. The PSPase
 CC polypeptides may be used as antigens in the production of antibodies
 CC against the PSPase polypeptides and in assays to identify modulators of
 CC PSPase expression and activity. The anti-PSPase antibodies and
 CC antagonists may also be used to down regulate expression and activity,
 CC the anti-PSPase antibodies may also be used as diagnostic agents for

CC detecting the presence of PSPase polypeptides in samples. Disorders that
 CC may be prevented, diagnosed and/or treated by the invention are:
 CC immune/autoimmune disorders (e.g. Wiscott-Aldrich syndrome,
 CC Chediak-Higashi syndrome, Hashimoto's thyroiditis and multiple
 CC sclerosis); inflammatory conditions (e.g. Crohn's disease, inflammatory
 CC bowel disease, appendicitis and rheumatoid arthritis); cellular
 CC proliferative disorders (e.g. lymphoma, lung and intestinal cancers); and
 CC cardiovascular disorders (e.g. Scimitar syndrome, Ebstein's anomaly and
 CC aneurysm).

XX SQ Sequence 1878 BP; 479 A; 499 C; 417 G; 483 T; 0 other;

Alignment Scores:

Pred. No.: 2,61e-230 Length: 1878
 Score: 2343.00 Matches: 449
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 22 Gaps: 0

US-09-941-831A-20 (1-449) x AAH78730 (1-1878)

QY 1 MetArgGluSerGlyTyrPlysLeuLeuAspProIleSerAspPheGlyArgMetGlyIle 20
 |||||
 Db 367 ATGAGGAAAGTGGATGGAACCTGATTGACCAATATCAGACTTTGGCGTATGGAATA 426
 |||||
 QY 21 ProAsnArgAsnTyrThrIleThrAspAlaAsnArgAsnTyrGluIleCysSerThrTyr 40
 |||||
 Db 427 CCCAACACAGAACTGGACCATACAGATGCCAACAGAACTATGAGATATGACGACCTTAC 486
 |||||
 QY 41 ProProGluIleValValProLysSerValThrLeuGlyThrValValGlySerSerLys 60
 |||||
 Db 487 CCTCTGAAATAGTGGTTCCTAAATCTGTTACCTGGGAACGGTGGTGGAAAGTTCAAG 546
 |||||
 QY 61 PheArgSerLysGluArgValProValLeuSerTyrLeuTyrLysGluAsnAlaAla 80
 |||||
 Db 547 TTCAGAGTAAAGAACGTGTCCCTGTCTCTCTACCTCTACAAAGAGAACAAATGCTGCC 606
 |||||
 QY 81 IleCysArgCysSerGlnProLeuSerGlyPheTyrThrArgCysValAspAspGluLeu 100
 |||||
 Db 607 ATTGCGCGCTGTAGCAGCCTCTCTCGATTTTACACTCGCTGTGTAGATGATGAGCTC 666
 |||||
 QY 101 LeuLeuGluAlaIleSerGlnThrAsnProGlySerGlnPheMetTyrValValAspThr 120
 |||||
 Db 667 TTGTTGGAGGCCATTAGCCAAACAAACCCAGGAGCCAGTTTATGTATGTTGTAGACACA 726
 |||||
 QY 121 ArgProLysIleTyrPhePheLeuValLeuLeuMetArgIleValLeuGlnLeuAlaLys 140
 |||||
 Db 727 AGACCAAAAGATCTGGCATTTCTTGTGCTCATATGAGAAATAGTTCTCCAATTAGCCAAG 786
 |||||
 QY 141 MetAsnLeuMetAspIleThrLysIlePheSerIleLeuGlnProAspLysGluGlu 160
 |||||
 Db 787 ATGAACCTCATGGACATCACCAGATCTTCTCCCTCTGACGCCGACAGGAGGAGGAG 846
 |||||
 QY 161 AspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyr 180
 |||||
 Db 847 GACATGTACACAGAGAGAGCAGGCTCTCAATCAAGCAGTGATGTATGACAACTCTTAT 906
 |||||
 QY 181 ThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGly 200
 |||||
 Db 907 ACTTTGGACCAGCTTTTGGCCAGGAGCGTTTACAACAGTTTTCATCAACAGCAGGAGTGC 966
 |||||
 QY 201 TrpGlyValProGlyThrProLeuArgLeuAlaIleSerTyrGlyHisLeuSerCysLeu 220
 |||||
 Db 967 TGGGGTGTTCCTGGGACACCTTGGCGTGGCTGCTTCTTATGGCCACTTGTAGCTGTTTG 1026
 |||||
 QY 221 GlnValLeuLeuAlaHisGlyValAspValAspSerLeuAspValLysAlaGlnThrPro 240
 |||||
 Db 1027 CAAGTCTCTTAGCCCTAGTGTGATGTGACCTTGGATGTCAAGGCACAGACGCCCA 1086
 |||||
 QY 241 LeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGly 260
 |||||
 Db 1087 CTTTTCACGTGCTGTCAGTCATGCCCATCTGGACTGTGTACGTGTGCTTTTGGAAAGCTGT 1146
 |||||

```
QY 261 AlaSerProGlyGlySerIleTyrAenAenCysSerProValIleuThrAlaalaargasp 280
Db 1147 GCCTCTCCCTGGGTAGCATCTACAACTGTTCTCCCGTGTCTACAGCTGCCGTGAT 1206
QY 281 GlyAlaValAlaIleLeuGlnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 300
Db 1207 GGTGCTGTGCTATCTCTGAGGAGCTCTTAGACCATGTTGTCAGAGGCCAACGTCAGCT 1266
QY 301 LysLeuProValTrpAlaSerIleAlaSerCysSerGlyProLeuTyrIleuAlaAla 320
Db 1267 AAACCTACCAAGTCTGGGCATCAACATAGCTTCTGTGGCCCTCTTATTTGGCCGCA 1326
QY 321 ValTyrGlyHisLeuAspCysPheArgLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 340
Db 1327 GTCTACGGGCACCTGGACCTGTTTCCGCCCTGCTTTTGTCTCCAGGGGCAGACCTGACTAC 1386
QY 341 AsnCysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluLe 360
Db 1387 AACTGCATGACCGAGGCGCTATTGGCTCGTGTCCCAAGACCCCGCACCTCTCTGAAATC 1446
QY 361 CysLeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuLeuLeuLeuLeuLeu 380
Db 1447 TGCCTCCATCATATTTGTGAGCAGAGTATATCCAGCTGTTAATCGATTTTGTGCTAAT 1506
QY 381 IleTyrLeuProSerLeuSerLeuAspLeuThrSerGlnAspAspGlyIleAlaLeu 400
Db 1507 ATCTACCTTCCATCTCTCTCCCTTGACCTGAGCTCAAGAGTATTAAGGCAATTCGATTG 1566
QY 401 LeuLeuGlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArg 420
Db 1567 CTGCTACAGCCGCGAGCCACTCCAGGTCACCTTCTATCATCAGGTCGCTTAGTCTCCGC 1626
QY 421 ArgAlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMet 440
Db 1627 AGAGCCTTGTGCCAGGCTGCCAGGCACCAAGCCATCAACAGCTGGATATTCTCCCATG 1686
QY 441 LeuLeuSerTyrLeuLysHisGlnLeu 449
Db 1687 TTGATTAGTACTTAACCAACCAACTG 1713

RESULT 3
AAS33372
ID AAS33372 standard; cDNA; 1878 BP.
XX
AC AAS33372;
XX
DT 04-DEC-2001 (first entry)
XX
DE DNA encoding human secreted protein, Seq ID No 331.
XX
KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW Gaucher's disease; neurological disease; cerebrovascular disorder;
KW thrombosis; wound healing; ss.
XX
OS Homo sapiens.
XX
PN WO20015326-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01347.
XX
PR 31-JAN-2000; 2000US-0179065.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
```

WPI: 2001-451931/48.
P-PSDB; AAU20663.

New nucleic acids and polypeptides, useful for diagnosing, preventing
or treating medical conditions -

Claim 1; SEQ ID No 331; 753pp; English.

The invention relates to novel isolated nucleic acid molecules (I)
encoding human secreted proteins (II). (I) and (II) are used to prevent,
treat or ameliorate a medical condition in e.g. humans, mice, rabbits, in
goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
the prevention, treatment and diagnosis of diseases associated with
inappropriate expression of secreted proteins. (I) and complementary
sequences may also be used as DNA probes in diagnostic assays (e.g.
polymerase chain reactions (PCR)) to detect and quantitate the presence
of similar nucleic acid sequences in samples, and so which patients may
be in need of restorative therapy. (II) may also be used as antigens in
the production of antibodies and in assays to identify modulators
(agonists and antagonists) of the expression and activity of the secreted
proteins. The anti-(II) antibodies and antagonists may also be used to
down regulate expression and activity of (II). The anti-(II) antibodies
may also be used as diagnostic agents for detecting the presence of (II)
in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The
disorders include for example: immune/autoimmune diseases (e.g. HIV
(human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
melanomas, neoplasms of the breast or liver, Sezary syndrome and
Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
angina and thrombosis), infections caused by bacteria, viruses and
fungi and ocular disorders (e.g. corneal infections). (I) and (II),
agonists, antagonists and antibodies can also be used to promote wound
healing, maintain organs and before transplantation, and support cell culture
of primary tissues. AAS33043-AAS33486 represent human secreted protein
coding sequences, PCR primers, and related sequences of the invention.
Note: the sequence data for this patent did not appear in the printed
specification but was obtained in electronic format directly from WIPO
at: ftp.wipo.int/pub/published_pct_sequences.

Sequence 1878 BP; 479 A; 499 C; 417 G; 483 T; 0 other;

Alignment Scores:
Pred. No.: 2,61e-230 Length: 1878
Score: 2343.00 Matches: 449
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-09-941-831A-20 (1-449) x AAS33372 (1-1878)

QY 1 MetArgGluSerGlyTyrIleValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 20
Db 367 ATGAGGGAAGTGGATGGAACTGATTGACCAATATCAGACTTTGGGGGTATGGGAATA 426
QY 21 ProAsnArgAsnTyrThrIleThrAspAlaAsnArgAsnTyrGluIleCysSerThrTyr 40
Db 427 CCCAACAGAACTGGACCATTAACAGATGCCAAGAAATATGAGATATGCAGACCTAC 486
QY 41 ProProGluIleValValProLysSerValThrLeuGlyThrValValGlySerSerLys 60
Db 487 CCTCTGAAATAGTGGTTCCTAAATCTGTACCTGGGAACGGTGTGGTGGATGTTCAAG 546
QY 61 PheArgSerLysGluArgValProValLeuSerTyrLeuTyrLysGluAsnAlaAla 80
Db 547 TTCAGAAGTAAAGACGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 606
QY 81 IleCysArgCysSerGlnProLeuSerGlyPheTyrThrArgCysValAspAspGluLeu 100
Db 607 ATTTGCCGCTGTAGCCAGCCTCTCTCTGGATTTTACTACTCGCTGTGTAGATGATGACTC 666

```
QY 101 LeuLeuGluAlaIleSerGlnThrAsnProGlySerGlnPheMetTyrValValAspThr 120
Db 667 TTGTTGGAGCCATTAGCCAAACAAACCCAGGAGCCAGTTTATGTATGTTGTAGACACA 726
QY 121 ArgProLysIleTrpHisPheLeuValLeuIleMetArgIleValLeuGlnLeuAlaLys 140
Db 727 AGACCAAGATCTGGCATTTCTTGTGCTCATTAATGAGAATAGTTCTCCAAATTAGCCAA 786
QY 141 MetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGlu 160
Db 787 ATGAACCTCATGGACATCAACAGATCTTCTCTCTCCAGCCGACAGAGAGAGGAG 846
QY 161 AspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValIleAspAsnSerTyr 180
Db 847 GACACTGCACAGAGAGAGAGAGGCTCTCAATCAAGCAGTGTATGACAAACGACTCCTAT 906
QY 181 ThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGly 200
Db 907 ACTTTGGACCACTTTTGGCCGAGGAGCGTTACAAACGTTTTCATCAACAGCAGGAGTGC 966
QY 201 TrpGlyValProGlyThrProLeuArgLeuAlaIleSerTyrGlyHisLeuSerCysLeu 220
Db 967 TGGGGTGTCTTGGGACACCTTGGCTTGGCTGTCTTCTTATGACCACCTTGAGCTGTTG 1026
QY 221 GlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValIleAspAlaGlnThrPro 240
Db 1027 CAAGTCTCTTAGCCCATGCTGATGTTGACAGCTTGATGTCAAGGCACAGACGCCA 1086
QY 241 LeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGly 260
Db 1087 CTTTTCACCTGCTCAGTCATGCCATCTGGACATCTGTGATCGTGTCTTGTGGAAGCTGT 1146
QY 261 AlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAsp 280
Db 1147 GCCTCTCTCGGTGGTAGCATCTACAAACATGTCTTCCCGCTGTACAGCTGCCCGTGTAT 1206
QY 281 GlyAlaValAlaIleLeuGlnLeuLeuAspHisGlyAlaGluAlaAsnValIleAla 300
Db 1207 GGTGCTGTGCTATCTCTGAGGAGCTCTTAGACCATGTTGACAGGCCAACGTCAAAGCT 1266
QY 301 LysLeuProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrIleuAlaAla 320
Db 1267 AAACCTACCACTCTGGGCATCAACATAGCTTCTATGTTCTGGCCCCCTCTATTGCGCGCA 1326
QY 321 ValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyr 340
Db 1327 GTCTACGGGCACCTGGACTGTTCGCGCTGCTTTTGTCTCCAGGGGACACCTGACTAC 1386
QY 341 AsnCysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIle 360
Db 1387 AACTGCACCTGACAGGGCCTATTGGCTCGTGTCCCAAGACCCCGCACCTCTCTTGAATC 1446
QY 361 CysLeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsn 380
Db 1447 TGCCTCCATCAATAATTGTGAGCAGAGTATATCCAGCTGTAATCAATTTGGTGTCTAAT 1506
QY 381 IleTyrLeuProSerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeu 400
Db 1507 ATCTACTCTTCCATCTCTCTCCCTTGACCTGACCTCAAGATGATAAAGGCATTGCTATG 1566
QY 401 LeuLeuGlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArg 420
Db 1567 CTGCTACAGCCCGAGCCACTCCAGGTCATCTTATCATCAGGTCGTTTGTAGTCTCCGC 1626
QY 421 ArgAlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProMet 440
Db 1627 AGAGCCTTGTGCGAGCTGCGCCAGCCACAAAGCCATCAACAGCTGGATATCTCTCCCATG 1686
QY 441 LeuIleSerTyrLeuLysHisGlnLeu 449
Db 1687 TTGATTAGTACTCTAAACACCAACTG 1713
```

RESULT 4

```
AAS33264/c
ID AAS33264 standard; cDNA; 2530 BP.
XX AC AAS33264;
XX DT 04-DEC-2001 (first entry)
XX DE DNA encoding human secreted protein, Seq ID No 223.
XX KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW rheumatoid arthritis; arteriosclerotic; cardiac; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW Gaucher's disease; neurological disease; cerebrovascular disorder;
KW thrombosis; wound healing; ss.
XX OS Homo sapiens.
XX PN WO200155326-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01347.
XX PR 31-JAN-2000; 2000US-0179065.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Barash SC, Ruben SM;
XX DR WPI; 2001-451931/48.
XX DR P-PSDB; AAU20555.
XX PT New nucleic acids and polypeptides, useful for diagnosing, preventing
XX PT or treating medical conditions -
XX PS Claim 1; SEQ ID No 223; 753pp; English.
XX CC The invention relates to novel isolated nucleic acid molecules (I)
XX CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
XX CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
XX CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
XX CC the prevention, treatment and diagnosis of diseases associated with
XX CC inappropriate expression of secreted proteins. (I) and complementary
XX CC sequences may also be used as DNA probes in diagnostic assays (e.g.
XX CC polymerase chain reactions (PCR)) to detect and quantitate the presence
XX CC of similar nucleic acid sequences in samples, and so which patients may
XX CC be in need of restorative therapy. (II) may also be used as antigens in
XX CC the production of antibodies and in assays to identify modulators
XX CC (agonists and antagonists) of the expression and activity of the secreted
XX CC proteins. The anti-(II) antibodies and antagonists may also be used to
XX CC down regulate expression and activity of (II). The anti-(II) antibodies
XX CC may also be used as diagnostic agents for detecting the presence of (II)
XX CC in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The
XX CC disorders include for example: immune/autoimmune diseases (e.g. HIV
XX CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
XX CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
XX CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
XX CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
XX CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
XX CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
XX CC angina and thrombosis), infections caused by bacteria, viruses and
XX CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
XX CC agonists, antagonists and antibodies can also be used to promote wound
XX CC healing, maintain organs before transplantation, and support cell culture
XX CC of primary tissues. AAS33043-AAS33486 represent human secreted protein
XX CC coding sequences, PCR primers, and related sequences of the invention.
XX CC Note: the sequence data for this patent did not appear in the printed
XX CC specification but was obtained in electronic format directly from WIPO
XX CC at: ftp.wipo.int/pub/published_pct_sequences.
XX CC Sequence 2530 BP; 630 A; 602 C; 675 G; 614 T; 9 other;
```

Alignment Scores:

Pred. No.: 6.9e-229 Length: 2530
 Score: 2331.00 Matches: 447
 Percent Similarity: 99.55% Conservatives: 0
 Best Local Similarity: 99.55% Mismatches: 2
 Query Match: 99.49% Indels: 0
 DB: 22 Gaps: 0

US-09-941-831A-20 (1-449) x AAS33264 (1-2530)

QY 1 MetArgGluSerGlyTrpLysLeuIleAspProIleSerAspPheGlyArgMetGlyIle 20
 DB 1575 ATGAGGGAAGTGGATGGAAATGATGACCCAAATATCAKACTTTGGCGGTATGGGAATA 1516
 QY 21 ProAsnArgAsnTrpThrIleThrAspAlaAsnArgAsnTyrGluIleCysSerThrTyr 40
 DB 1515 CCCAAGAGAACTGGACCAATAMCAGATGCCAAGCAAGCAATATGAGATATGCAGACCTAC 1456
 QY 41 ProProGluIleValValProLysSerValThrLeuGlyThrValValGlySerSerLys 60
 DB 1455 CCTCTCGAATAGTGGTCTCTAATCTGTACCTTGGGAACGGTGGTGGAGTTCAGAG 1396
 QY 61 PheArgSerLysGluArgValProValLeuSerTyrLeuTyrLysGluAsnAlaAla 80
 DB 1395 TTCAGAAAGTAAGAACGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1336
 QY 81 IleCysArgCysSerGlnProLeuSerGlyPheTyrThrArgCysValAspAspGluLeu 100
 DB 1335 ATTTCGCGCTGTAGCCAGCGCTCTCTGGATTTTACACTGGCTGTAGATGATGAGCTC 1276
 QY 101 LeuLeuGluAlaIleSerGlnThrAsnProGlySerGlnPheMetTyrValValAspThr 120
 DB 1275 TTGTTGGAGCCATTAGCCAAACAAACCCAGGAGCCAGTTTATGATGTTGTAGACACA 1216
 QY 121 ArgProLysIleTrpHisPheLeuValLeuIleMetArgIleValLeuGlnLeuAlaLys 140
 DB 1215 AGACCAAGATCTGGCATTTCTTGTCTCATATAGAAATAGTCTCTCAATTTAGCCAAAG 1156
 QY 141 MetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGlu 160
 DB 1155 ATGAACCTCATGGACATCAACCAAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1096
 QY 161 AspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyr 180
 DB 1095 GACACTGACACAGAGAGAGAGAGGCTCTCAATCAAGCAGTGTATGACACGACTCTCTAT 1036
 QY 181 ThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGly 200
 DB 1035 ACTTTGGACACAGCTTTTGGCCAGGAGCGTTACAAACGTTTCATCAACAGCAGGAGTGC 976
 QY 201 TrpGlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeu 220
 DB 975 TGGGGTGTCTCTGGGACACCTTGGCGCTTGGCTCTCTTATGGCCACTTGAGCTGTTG 916
 QY 221 GlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrPro 240
 DB 915 CAACTCTCTTAGCCCATGGT 856
 QY 241 LeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGly 260
 DB 855 CTTTTCACTGCTGCTAGTCATGCCATCTGGACTGTGTGCTGTGTGTGTGTGTGTGTGTGT 796
 QY 261 AlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAsp 280
 DB 795 GCCTCTCTCTGGTGTAGCATCTACAACTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 736
 QY 281 GlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnValLysAla 300
 DB 735 GGTGCTGTGTCTATCTCTGAGGAGCTCTTAGACCATGTGTGAGAGCCCAACGTCAAGCT 676
 QY 301 LysLeuProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAla 320

DB 675 AAACCTACCAGTCTGGGCATCAACATAGCTTCTGTCCTCTCTATTTGGCCGCA 616
 QY 321 ValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyr 340
 DB 615 GTCACGGCACCTGGACGTGTTTCCGCTGCTTTTGTCTCCAGGGCAGACCTGACTAC 556
 QY 341 AsnCysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIle 360
 DB 555 AACTGCACCTGACACGGGCTATTGGCTGTGTCCCAAGACCCCGCACCTCTCTTGAATC 496
 QY 361 CysLeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuLeuAspPheGlyAlaAsn 380
 DB 495 TGCCTCCATCATATTTGTGAGCCAGATATCCAGCTGTTAATCGATTTGGTGTAAAT 436
 QY 381 IleTyrLeuProSerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeu 400
 DB 435 ATCTACCTTCCATCT 376
 QY 401 LeuLeuGlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArg 420
 DB 375 CTGCTACAGGCCCGAGCCACTCCACGGTCACCTTCTATCAGAGTCCGTTTAGTCGTCGC 316
 QY 421 ArgAlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMet 440
 DB 315 AGAGCCTTTGTGCAGGCTGGCCAGCCACCAAGCCATCAACAGCTGGATATTCCTCCCATG 256
 QY 441 LeuIleSerTyrLeuLysHisGlnLeu 449
 DB 255 TTGATTAGTACTACCTAAACACCAACTG 229
 RESULT 5
 ID ABK11107 standard; DNA; 1222 BP.
 AC ABK11107;
 DT XX
 DT XX
 XX 05-JUN-2002 (first entry)
 DE DNA encoding human NOV5b protein, homologue of ankyrin-repeat proteins.
 XX Human; NOVX-associated disorder; developmental disorder; blood disorder;
 KW endocrine disorder; vascular disease; gastrointestinal disorder; cancer;
 KW respiratory disorder; inflammatory disorder; reproductive disorder;
 KW neurodegenerative disorder; autoimmune disorder; infectious disease;
 KW cardiovascular disorder; cell signal processing; ankyrin-repeat protein;
 KW metabolic pathway modulation; NOV5b; gene; ds.
 XX
 OS Homo sapiens.
 FH Key Location/Qualifiers
 CDS 6..1088
 FT /*tag= a
 FT /product= "NOV5b"
 XX
 PN WO200206329-A2.
 XX
 PD 24-JAN-2002.
 XX
 PF 18-JUL-2001; 2001WO-US22709.
 XX
 PR 18-JUL-2000; 2000US-218870P.
 PR 18-JUL-2000; 2000US-218875P.
 PR 18-JUL-2000; 2000US-218901P.
 PR 24-JUL-2000; 2000US-220273P.
 PR 26-JUL-2000; 2000US-220912P.
 PR 27-JUL-2000; 2000US-221233P.
 PR 28-JUL-2000; 2000US-221650P.
 XX
 PA (CURA-) CUPAGEN CORP.
 XX
 PI Rastelli L, Shinkets RA, Zerhusen B, Malyankar UM, Padigaru M;
 XX WPI; 2002-179781/23.
 DR

DR P-PSDB; AAU77411.

XX Novel cytoplasmic, nuclear, membrane bound and secreted NOVX

PT polypeptides, useful for treating developmental disorders, endocrine

PT disorders, vascular disorders, infectious diseases and

PT neurodegenerative disorders -

XX

PS Claim 9; Page 39; 178pp; English.

XX

CC The present invention relates to the isolation of novel human

CC polypeptides referred to as NOV1, NOV2, NOV3, NOV4a, NOV4b, NOV5a,

CC NOV5b, NOV6 AND NOV7, and the polynucleotide sequences encoding them.

CC The NOVX polypeptides are related to NOPB, cadherin, interferon

CC alpha-13, ADAM, ankyrin repeat-containing, transpanin or semaphorin

CC polypeptides. The sequences of the invention are useful for identifying

CC an agent (a cellular receptor or downstream effector) that binds to

CC the NOVX polypeptide, or an agent that modulates its expression or

CC activity. They are useful for treating or preventing NOVX-associated

CC disorders such as developmental disorders, endocrine disorders, vascular

CC diseases, gastrointestinal disorders, respiratory disorders, inflammatory

CC disorders, blood disorders, reproductive disorders, neurodegenerative

CC disorders, autoimmune and immune disorders, infectious diseases,

CC cardiovascular disorders, cancers, and other disorders related to cell

CC signal processing and metabolic pathway modulation. The present sequence

CC encodes the human NOV5b protein.

XX

SX Sequence 1222 BP; 276 A; 352 C; 278 G; 316 T; 0 other;

Alignment Scores:

Pred. No.: 1,58e-163 Length: 1222

Score: 1690.00 Matches: 326

Percent Similarity: 99.70% Conservative: 1

Best Local Similarity: 99.39% Mismatches: 1

Query Match: 72.13% Indels: 0

DB: 24 Gaps: 0

US-09-941-831A-20 (1-449) x ABK11107 (1-1222)

QY 122 ProLysLeuThrPheLeuValLeuMetArgIleValLeuGlnLeuAlaLysMet 141

DB 102 CCACAGATCTGGCATTCTCTTGGTCTAATGAGAAATAGTCTCCAAATGACCAAGATG 161

QY 142 AsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGluAsp 161

DB 162 AACCTCATGGACATACCAAGATCTCTCCCTCCAGCCGACAGAGAGAGAGAC 221

QY 162 ThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThr 181

DB 222 ACTGACACAG 281

QY 182 LeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGlyTrp 201

DB 282 TTGGACACAGCTTTGGCCAGGAGCGTTTACAAACGTTTCATCAACAGCAGAGAGTGGCTGG 341

QY 202 GlyValProGlyThrProLeuArgLeuAlaSerTyrGlyHisLeuSerCysLeuGln 221

DB 342 GGTGTTCTCGGACACCCCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTT 401

QY 222 ValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeu 241

DB 402 GTCTCTTAGCCATGGTCTGATGTTGACAGCTTGGATGTCAAGGCACAGACGCCACTT 461

QY 242 PheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAla 261

DB 462 TTCACCTGCTGTCAGTCATGCCATCTCGGACTGTGTACGTGTGCTTTTGGAACTGGTGCC 521

QY 262 SerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGly 281

DB 522 TCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 581

QY 282 AlaValAlaIleLeuGlnLeuLeuAspHisGlyAlaGluAlaAsnValLysAlaLys 301

DB 582 GCTGTGTCTATCTCTGACAGAGCTCTTAGACCATGGTGGTGGTGGTGGTGGTGGTGGTGGT 641

QY 302 LeuProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaVal 321

DB 642 CTACAGCTCTGGCATCAAAATAGCTTCATGTTCTGGCCCCCTTATTGGCCGAGTC 701

QY 322 TyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsn 341

DB 702 TACGGGACCTTGGACTGCTTTCCGCTGCTTTTGGCTCCACGGGGCAGACCTGACTACAAC 761

QY 342 CysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluLeuCys 361

DB 762 TGCACTGACCGAGGCTTATGGCTGCTGTCACAGACCCCGACCCCTCTTGAATCTGC 821

QY 362 LeuHisHisAsnCysGluProGlyTyrIleGlnLeuLeuLeuAspPheGlyAlaAsnIle 381

DB 822 CTCATCATATATTGAGCCAGAGATATACGCTGTTAATCGATTTTGGTGTCTAATATC 881

QY 382 TyrLeuProSerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeuLeu 401

DB 882 TACCTTCCATCTCTCTCCCTTGGCTGACCTCACAGATGATAAAGGCAATTCATTTGCTG 941

QY 402 LeuGlnAlaArgAlaThrProArgSerLeuLeuSerGlnValAlaTqLeuValValArg 421

DB 942 CTACAGGCGGAGCCACCTCCACGGTCACTTCTATCACAGGTCCTGTTAGTCTCCGCGA 1001

QY 422 AlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMetLeu 441

DB 1002 GCCTTGTCCAGGCTGGCCAGCCACAGCCATCAACGAGTGGATATTTCTTCCCATGTTG 1061

QY 442 IleSerTyrLeuLysHisGlnLeu 449

DB 1062 ATTAGCTACCTAAAACACCAACTG 1085

RESULT 6

ABZ11533

ID ABZ11533 standard; cDNA; 1127 BP.

XX

AC ABZ11533;

XX

DT 20-JAN-2003 (first entry)

XX

DE Human polynucleotide SEQ ID NO 415.

XX

KW Human; genome mapping; gene therapy; food supplement; virus; fungus;

KW cell-proliferative disorder; neurodegenerative disease; bacterial;

KW Parkinson's disease; Alzheimer's disease; autoimmune disease;

KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;

KW arthritis; cytostatic; immunomodulator; neoplastic; dermatological;

KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;

KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;

KW antiarthritic; gene; ss.

XX

OS Homo sapiens.

XX

XX WO200270539-A2.

PN

XX

PD 12-SEP-2002.

XX

XX 05-MAR-2002; 2002WO-US05095.

PF

XX

PR 05-MAR-2001; 2001US-0799451.

XX

XX (HYSE-) HYSEQ INC.

PA

XX

PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;

PI Wehrman T, Wang J, Wang D, Drmanac RT;

XX

DR WPI; 2002-759812/82.

DR P-PSDB; ABP69316.

XX

PT New polynucleotides comprising sequences assembled from expressed

PT sequence tags (ESTs), useful for treating cell-proliferative,

PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
 PT platelet or coagulation disorders -

XX Claim 1; SEQ ID NO 415; 1012pp + Sequence Listing; English.

XX The invention relates to an isolated polynucleotide (I) comprising a
 CC nucleotide sequence selected from any of 948 sequences
 CC (ABZ1119-ABZ12066) or their mature protein coding portion, active domain
 CC coding protein or complementary sequences. The polynucleotides are useful
 CC for identifying expressed genes or for physical mapping of human genome.
 CC The encoded polypeptides (ABP68902-ABP69849) are useful as molecular
 CC weight markers, as a food supplement, for generating antibodies, in
 CC medical imaging, screening and diagnostic assays and for treating
 CC cell-proliferative disorders (cancer), neurodegenerative diseases
 CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
 CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
 CC disorders, platelet or coagulation disorders, wound, burns, incision,
 CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
 CC parasitic), arthritis, etc.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1127 BP; 244 A; 329 C; 281 G; 273 T; 0 other;

Alignment Scores:

Pred. No.: 1 49e-162 Length: 1127
 Score: 1680.00 Matches: 324
 Percent Similarity: 99.69% Conservative: 1
 Best Local Similarity: 99.39% Mismatches: 1
 Query Match: 71.70% Indels: 0
 DB: 24 Gaps: 0

US-09-941-831A-20 (1-449) x ABZ11533 (1-1127)

QY 124 ileftrPhisPheLeuValLeuLeuMetArgileValLeuGlnLeuAlaLysMetAsnLeu 143
 DB 135 ATCTGGCATTTCTCTGTCTATAATGAGATAGTTCTCCATTAGCCAGATGACCTC 194
 QY 144 MetAspIleThrIysIlePheSerLeuLeuGlnProAspIysGluGluAspThrAsp 163
 DB 195 ATGGACATCACCAGATCTTCTCCCTCTGCGAGCCGACAGAGGAGGAGGACACTGAC 254
 QY 164 ThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThrLeuAsp 183
 DB 255 ACAGAGGAGAGGAGGCTCTCATCAAGCAGGTGTATGACAACTCTCTATCTTTGGAC 314
 QY 184 GlnLeuLeuArgGlnGluArgTyrIysArgPheIleAsnSerArgSerGlyTrpGlyVal 203
 DB 315 CAGCTTTTGGCCAGAGGCGTTTACAAACGTTTCATCAACAGCAGGAGTGGCTGGGTGT 374
 QY 204 ProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeu 223
 DB 375 CTGGGACACCTTGGCTTGGCTGCTCTTATGGCCACTTGCAGCTGTTTGGAAAGTCCTC 434
 QY 224 LeuAlaHisGlyValAspValAspSerLeuAspValIysAlaGlnThrProLeuPheThr 243
 DB 435 TTAGCCCATGGTGTGATGTGACCTTGGATGTGATGTCAGGCACAGCCACTTTTCACT 494
 QY 244 AlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerPro 263
 DB 495 GCTGTTCAGTCATGGCCATCTGGACTGTGTACGTGTCTTTTGGAAAGCTGTGCTCTCTC 554
 QY 264 GlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaVal 283
 DB 555 GGTGTGTAGATCTTACAACTGTTCTCCGCTGCTCACAGCTGCCGCTGATGCTGTGT 614
 QY 284 AlaIleLeuGlnLeuLeuAspHisGlyAlaGluAlaAsnValIysAlaLysLeuPro 303
 DB 615 GCTATCTCGAGAGCTCTTAGACCATGGTGCAGAGGCCAACGTCAAGCTTAACCTACCA 674
 QY 304 ValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAlaValTyrGly 323

DB 675 GTCTGGCATCAACATAGCTTTCTATTTCTGGCCCCCTCTATTTGGCCGCGAGTCTACGGG 734
 QY 324 HisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsnCysThr 343
 DB 735 CACCTGGACTGTTTCCGCTGCTTTTGCTCCACGGGGCAGACCTGACTACAACTGCAC 794
 QY 344 AspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIleCysLeuHis 363
 DB 795 GACCAAGGCTATTGGCTGCTGCCAGACCCCGACCTCTCTTGAATCTGCCTCCAT 854
 QY 364 HisAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsnIleTyrLeu 383
 DB 855 CATAATTGTAGCCAGAGATATATCCAGCTGTTAAATCGATTTTGGTCTAATATCTACCT 914
 QY 384 ProSerLeuSerLeuAspIleThrSerGlnAspAspIysGlyIleAlaLeuLeuGln 403
 DB 915 CCATCTCTCTCCCTTGACCTGACCTCACAAAGATATAAGGCATTGCATTGCTGTACAG 974
 QY 404 AlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgAlaLeu 423
 DB 975 GCGCGAGCCACTCCACGGTCACTTCTATCACAGGTCCGTTAGTCGTCGCGAGGCTTG 1034
 QY 424 CysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMetLeuIleSer 443
 DB 1035 TGCCAGGCTGGCCAGCCACAGCCATCAACAGCTGGATATCTCTCCCATGTTGATTAA 1094
 QY 444 TyrLeuLysHisGlnLeu 449
 DB 1095 TACCTAAAACACCACCTG 1112
 XX ABK11106
 ID ABK11106 standard; DNA; 1069 BP.
 XX AC ABK11106;
 DT XX
 XX 05-JUN-2002 (first entry)
 DE DNA encoding human NOV5a protein, homologue of ankyrin-repeat proteins.
 KW Human; NOVX-associated disorder; developmental disorder; blood disorder;
 KW endocrine disorder; vascular disease; gastrointestinal disorder; cancer;
 KW respiratory disorder; inflammatory disorder; reproductive disorder;
 KW neurodegenerative disorder; autoimmune disorder; infectious disease;
 KW cardiovascular disorder; cell signal processing; ankyrin-repeat protein;
 KW metabolic pathway modulation; NOV5a; gene; ds.
 XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 CDS 6..1061
 FT /*tag= a
 FT /product= "NOV5a"
 XX
 XX WO200206329-A2.
 PD 24-JAN-2002.
 XX
 PF 18-JUL-2001; 2001WO-US22709.
 XX
 PR 18-JUL-2000; 2000US-218870P.
 PR 18-JUL-2000; 2000US-218875P.
 PR 18-JUL-2000; 2000US-218901P.
 PR 24-JUL-2000; 2000US-220273P.
 PR 26-JUL-2000; 2000US-220912P.
 PR 27-JUL-2000; 2000US-221233P.
 PR 28-JUL-2000; 2000US-221650P.
 XX
 XX (CURA-) CUPAGEN CORP.
 PA
 PI Rastelli L, Shinkets RA, Zerhusen B, Malyankar UM, Padigar M;
 XX WPI; 2002-179781/23.
 XX

DR P-PSDB; AAU77410.

XX Novel cytoplasmic, nuclear, membrane bound and secreted NOVX
PT polypeptides, useful for treating developmental disorders, endocrine
PT disorders, vascular disorders, infectious diseases and
PT neurodegenerative disorders -

XX Claim 9; Page 38; 178pp; English.

XX The present invention relates to the isolation of novel human
CC polypeptides referred to as NOV1, NOV2, NOV3, NOV4a, NOV4b, NOV5a,
CC NOV5b, NOV6 AND NOV7, and the polynucleotide sequences encoding them.
CC The NOVX polypeptides are related to NOPE, cadherin, interferon
CC alpha-13, ADAM, ankyrin repeat-containing, transpanin or semaphorin
CC polypeptides. The sequences of the invention are useful for identifying
CC an agent (a cellular receptor or downstream effector) that binds to
CC the NOVX polypeptide, or an agent that modulates its expression or
CC activity. They are useful for treating or preventing NOVX-associated
CC disorders such as developmental disorders, endocrine disorders, vascular
CC diseases, gastrointestinal disorders, respiratory disorders, inflammatory
CC disorders, blood disorders, reproductive disorders, neurodegenerative
CC disorders, autoimmune and immune disorders, infectious diseases,
CC cardiovascular disorders, cancers, and other disorders related to cell
CC signal processing and metabolic pathway modulation. The present sequence
CC encodes the human NOV5a protein.

XX Sequence 1069 BP; 237 A; 298 C; 257 G; 277 T; 0 other;

Alignment Scores:

Pred. No.: 2,38e-157 Length: 1069
Score: 1629.00 Matches: 318
Percent Similarity: 97.55% Conservative: 0
Best Local Similarity: 97.55% Mismatches: 0
Query Match: 69.53% Indels: 8
DB: 24 Gaps: 1

US-09-941-831A-20 (1-449) x ABK1106 (1-1069)

QY 132 MetArgIleValLeuGlnLeuAlaLysMetAsnLeuMetAspIleThrLysIlePheSer 151
DB 81 ATGAGAAATAGTTCTCCAAATAGGCAAGATGAACCTCATGGACATCAACCAAGATCTTCTCC 140
QY 152 LeuLeuGlnProAspLysGluGluAspThrAspThrGluLysGlnAlaLeuAsn 171
DB 141 CTCCTGCAGCCGACAAAGGAGGAGGACACTGACACAGAGAGAGAGAGGCTCTCAAT 200
QY 172 GlnAlaValThrAspAsnAspSerThrLeuAspGlnLeuArgGlnGluArgTyr 191
DB 201 CAAGCAGTGTATGACAACTCTATATCTTTGGACAGCTTTTGGCCAGGAGCGTTAC 260
QY 192 LysArgPheIleAsnSerArgSerGlyTyrGlyValProGlyThrProLeuArgLeuAla 211
DB 261 AAACGTTTTCATCAACAGCAGGAGTGCTGGGGTGTTCCTGGGACACCTTGGCTTGGCT 320
QY 212 AlaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAlaHisGlyAlaAspValAsp 231
DB 321 GCTTCTTATGGCCACTTGAGCTGTTTGAAGTCTCTTAGCCCATGGTCTGATGTGAC 380
QY 232 SerLeuAspValLysAlaGlnThrProLeuPheThrAlaValSerHisGlyHisLeuAsp 251
DB 381 AGCTTGGATGTCAAGGCACAGACGCCACTTTTCACTGCTGTGAGTCTATGGCCATCTGGAC 440
QY 252 CysValArgValLeuLeuGlnAlaGlyAlaSerProGlyGlySerIleTyrAsnAsnCys 271
DB 441 TGTGTACGTGTGCTTTTGAAGCTGGTGGCTCTCTCTGGTGGTGGTATCTACAACTGT 500
QY 272 SerProValLeuThrAlaAlaArgAspGlyAlaValAlaIleLeuGlnGluLeuLeuAsp 291
DB 501 TCTCCCGTCTACACGCTGCCGTGATGGTGTGCTATCTCTGAGGAGCTCTAGAC 560
QY 292 HisGlyAlaGlnAlaAsnValLysAlaLysLeuProValTrpAlaSerAsnIleAlaSer 311
DB 561 CATGGTGCAGAGGCCAACGCTCAAGCTAAAGTAACTACCACTCTGGGACATCAACATAGCTTCA 620

QY 312 CysSerGlyProLeuTyrLeuAlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeu 331
DB 621 TGTTCCTGGCCCTCTATTTGGCCGAGTCTACGGGACCTGGACTGTTTCCGCTGCTT 680
QY 332 LeuLeuHisGlyAlaAspProAspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgVal 351
DB 681 TTGCTCCACGGGGCAGACCTGACTTACAACTGCACTGACCCAGGGCTATTGGCTCGTGC 740
QY 352 ProArgProArgThrLeuLeuGluIleCysLeuHisAsnCysGluProGluTyrIle 371
DB 741 CCAGACCCCGCACCCTCTTGAATCTGCTCCATCATTAATTTGAGCCAGATATATC 800
QY 372 GlnLeuLeuLeuAspPheGlyAlaAsnIleTyrLeuProSerLeuSerLeuAspLeuThr 391
DB 801 CAGCTGTTAATGATTTGGTGTCTATATCTACCTTCCATCTCTCTCCCTGACCTGACC 860
QY 392 SerGlnAspAspLysGlyIleAlaLeuLeuGlnAlaArg----- 405
DB 861 TCACAAGATGATAAAGGCATTGCTTGTCTGCTACAGGCCGAGGTGAGCTGTTTCTTCTT 920
QY 406 -----AlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgAlaLeu 423
DB 921 GCTGTAGCCACTCCACGGTCACTTCTATCACAAGGTCCTTTAGTCTCGCAGAGCCTTG 980
QY 424 CysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMetLeuIleSer 443
DB 981 TGCCAGGCTGGCCAGCCACAGCCATCAACAGCTGGATATTCCTCCCATGTTGATTAGC 1040
QY 444 TyrLeuLysHisGlnLeu 449
DB 1041 TACCTAAAACACCAACTG 1058
RESULT 8
AAS77257/c
ID AAS77257 standard; cDNA; 8214 BP.
XX
AC AAS77257;
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #13061.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR P-PSDB; ABG13070.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 13061; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 8214 BP; 2876 A; 1791 C; 1749 G; 1798 T; 0 other;

Alignment Scores:

Pred. No.: 8,76e-58 Length: 8214
 Score: 672.00 Matches: 155
 Percent Similarity: 57.96% Conservative: 27
 Best Local Similarity: 49.36% Mismatches: 50
 Query Match: 28.68% Indels: 82
 DB: 23 Gaps: 10

US-09-941-831A-20 (1-449) x AAS77257 (1-8214)

QY 1 MetArgGluSerGlyTrpIysLeuIleAspProIleSerAspPheGlyArgMetGlyIle 20
 DB 1879 ATGGGGAAGTGGATGGAACTGATGGACCAATATCAGACTTTGGCGGTATGGGAATA 1820
 QY 21 ProAsnArgAsnTrpThrIleThrAspAlaAsnArgAsnTrpGluIleCysSerThrTrp 40
 DB 1819 CCCACAGAACTGGACCAATACAGATGCCACAGAACTATGAGATATGCAGCACTAC 1760
 QY 41 ProProGluIleValValProIysSerValThrLeuGlyThrValValGlySerSerIys 60
 DB 1759 CCTCTCGAATAGTGGTTCTTAATCTGTTCCTTGGGAACGGTGGTGGAGTTCAAAG 1700
 QY 61 PheArgSerIysGluArgValProValLeuSerTrpLeuTrpIysGluAsnAsnAlaA 80
 DB 1699 TTCAGAAAGTAAAGAACGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1640
 QY 81 IleCysArgCysSerGlnProLeuSerGlyPheTrpThrArgCysValAspAspGluLeu 100
 DB 1639 ATTTGGCGGTAGCCAGGCTCTCTCTGATTTTACACTCGCTGTGTGATGATGAGCTC 1580
 QY 101 LeuLeuGluAlaIleSerGlnThrAsnProGlySerGlnPheMetTrpValValAspThr 120
 DB 1579 TTGTTGGAGGCCATTAGCCAAACAAACCAGGAGCCAGTTTATGTATGTGTAGACACA 1520
 QY 121 ArgProIysIle----- 124
 DB 1519 AGACCAAGTTGAATGCCATGCCAACCGAGCAGCTGGGAAGGGGTATGAAATGAAGAC 1460
 QY 125 -----TrpHisPheLeu-----ValLeuIleMetArgIle 134
 DB 1459 AACTATGCCAACATTCGTTTCAGATTCATGGGCATTTGAGAACATCCATGATTCGGAGC 1400
 QY 135 ValLeuGlnLeuAlaIysMetAsnLeuMetAspIleThrLysIlePheSerLeuGln 154
 DB 1399 AGTCTGCAG-----AAACTCTTGGAGTTGTGCAATTGAAACT 1361
 QY 155 ProAspIysGluGluAspThrAspThrGluGluIysGlnAlaLeuAsn-----Gln 172
 DB 1360 CCAACAAATGAGTGAATTTCTAGCGCCCTGGAGAGCTCAGGGTGGTTAAGACACATATA 1301
 QY 173 AlaValTyAspAsnAspSerTrpThrLeuAspGlnLeuLeuArgGlnGluArgTrpIys 192

DB 1300 GCTATTATGATGCTGGAATTTTC---ATTACAAAGGCAGTGAAGGTAGAAAAGGCCAGT 1244
 QY 193 ArgPheIleAsnSerArgSerGlyTrpGlyValProGlyThrProLeuArgLeuAlaA 212
 DB 1243 GTCTTAGTCCCATTTGTTCTGATGATGGGACCCGACACAGCAAGTCTGCTCAGTGGCTAGC 1184
 QY 212 ----- 212
 DB 1183 ATCCTCTAGATCCATTTTATAGGACATTCAAAGGACTCATGAAAGGTGGTATCGTTGT 1124
 QY 213 -----SerTrp-----GlyHisLeuSerCysLeuGlnValLeuLeu 224
 DB 1123 GAGCCAGGTTGTAATTATATATCAGAAAGGTGTGCCACCTC----- 1085
 QY 225 AlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPheThrAla 244
 DB 1084 -----GATGGGACTCTAAAGAAGTG-----TCCCTATCTTCACCCAG 1046
 QY 245 ValSerHisGlyHisLeuAspCysValArgValLeuLeuGlu 258
 DB 1045 -----TTCTAGACTGTATCTGGCAATTAATGGAA 1016
 RESULT 9
 ABN60012
 ID ABN60012 standard; cDNA; 1345 BP.
 XX
 AC ABN60012;
 XX
 DT 28-JUN-2002 (first entry)
 XX
 DE Novel human coding sequence SEQ ID NO: 423.
 XX
 KW Human; antianemic; vulnary; antiinflammatory; immunomodulator;
 KW antifertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
 KW neuroprotective; antiparkinsonian; protein therapy; EST;
 KW expressed sequence tag; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200222660-A2.
 XX
 PD 21-MAR-2002.
 XX
 PF 10-SEP-2001; 2001WO-US26015.
 XX
 PR 11-SEP-2000; 2000US-0659671.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
 XX
 DR WPI; 2002-292408/33.
 DR P-PSDB; ABB97599.
 XX
 PT An isolated polynucleotide for treating diseases associated with its
 PT encoded polypeptide such as cancer and multiple sclerosis -
 XX
 PS Claim 1; SEQ ID NO 423; 509pp; English.
 XX
 CC The present invention provides the protein and coding sequences of 444
 CC novel human proteins. These were isolated from expressed sequences tags
 CC (ESTs). They can be used to stimulate cell growth, to regulate
 CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat
 CC multiple sclerosis, to regulate activin or inhibit e.g. to treat
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
 CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
 CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
 CC Parkinson's disease. The present sequence is a coding sequence of the
 CC invention.

119

Percent Similarity: 57.89% Conservative: 57
Best Local Similarity: 39.14% Mismatches: 102
Query Match: 21.30% Indels: 26
DB: 22 Gaps: 9

US-09-941-831A-20 (1-449) x AAS33071 (1-1266)

QY 166 GluLysGlnAlaLeuAenGlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeu 185
DB 1140 GAGGACACGAGGCTCCATGATGAGCTTACGTCGGGAGCTCCAGACCTCCAGAGCCTA 1081
QY 186 LeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGlyTrp-----Gly 202
DB 1080 TTGCAAGAGGAGAGCTACCGAGCCGATCAACAGAGAGTCTGTCTGGTGTCTGGCTGG 1021
QY 203 ValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnVal 222
DB 1020 CTCCTCTGCACACCGTTCGGAATCGCGGCCACTGCAGGCCATGGAGCTGTGTGGACTTC 961
QY 223 LeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPhe 242
DB 960 CTCATCCGAGAGGGCCGAGGTGATCTGGTGAGCTAAAGGACAGAGCGCCCTGTAT 901
QY 243 ThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSer 262
DB 900 GTGGCTGTGTGAACGGCACCTAGAGAGTACCAGATCTCTCTCGAAGCTGGCGCGAC 841
QY 263 ProGlyGlySerIleTyrAsnLysCysSerProValLeuThrAlaAlaArgAspGlyAla 282
DB 840 CCCAACGGAAGCGGCGGACCATCCAGCACACCCCTGTCTACACGCCCTCTCGCGCGCG 781
QY 283 ValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeu 302
DB 780 GCAGACATCTGAGGCCCTCATCAGGTAGCGGGCTGATGTGACGTCAACACACCTG 721
QY 303 -----ProValThrAlaSerAsnIleAlaSer-----CysSerGly 314
DB 720 ACTCTGTATGTCAGGCTCGATCTCCCGCGGCTCACCTCTCTGTGTGTCTGCTG 667
QY 315 ProLeuTyrLeuAlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHis 334
DB 666 CCTTTGTATCATCGCCAGCCCTACCAACACCTCCAGTGTCTCGGCTGTCTCTCTGGCT 607
QY 335 GlyAlaAspProAspTyrAsnCys-----ThrAspGlnGlyLeuLeuAlaArg 350
DB 606 GCGCGAACCCTGACTTCACTGCAATGTCTGTCTCAACACAGAGATTTACAGG---- 550
QY 351 ValProArgProArgThrLeuLeuGluIleCysLeuHisHisAsnCysGluProGluTyr 370
DB 549 ---GGCTCCCTGGGTGGCTCATGGATGCTGTCTCGCCACCGCTGTGAGGAGCCTTC 493
QY 371 IleGlnLeuLeuLeuAspPheGlyAlaAsnIleTyrLeu-----ProSerLeuSer 387
DB 492 GTGAGCTCTGTGTAGATTTGAGGCAACCTCTAGTGAAGTGGGAATCGCTGGGC 433
QY 388 LeuAspLeuThrSerGln-----AspAspLysGlyIleAlaLeuLeuGlnAla 404
DB 432 CCAGAGTCAGAGGAGAGAGAAAGTGGACCTTGAGGCCCTTGAGGTCTTTAAAGAGGCC 373
QY 405 ArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgArgAlaLeuCys 424
DB 372 AGAAGTGTCCAGACACCTTGTGTCTGTCTGCGGTGTGCTGTGAGAAGAGCTCTTGCC 313
QY 425 GlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMetLeuLeuSerTyr 444
DB 312 AAA---CACCGGCTTCATCTGATTCCTCGCTCTGCCAGACCCCATAAAGATT 256
QY 445 LeuLysHisGln 448
DB 255 CTACTCCATGAG 244

RESULT 11
ABV23231

ID ABV23231 standard; cDNA; 1433 BP.
XX
AC ABV23231;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 23222.
DE
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
OS
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JB;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 4184; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 1433 BP; 280 A; 377 C; 410 G; 336 T; 30 other;

Alignment Scores:
Pred. No.: 3,96e-41 Length: 1433
Score: 499.00 Matches: 119
Percent Similarity: 57.89% Conservative: 57
Best Local Similarity: 39.14% Mismatches: 102
Query Match: 21.30% Indels: 26
DB: 23 Gaps: 9

US-09-941-831A-20 (1-449) x ABV23231 (1-1433)
QY 166 GluLysGlnAlaLeuAenGlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeu 185
DB 109 GAGGACACGAGGCTCCATGATGAGCTTACGTCGGGAGCTCCAGACCTCCAGAGCCTA 168
QY 186 LeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGlyTrp-----Gly 202
DB 169 TTGCAAGAGGAGAGCTACCGAGCCGATCAACAGAGAGTCTGTCTGTGTGTGTGTGTG 228

```
QY 203 ValProGlyThrProLeuArgLeuAlaAAsrTyrGlyHisLeuSerCysLeuGlnVal 222
DB 229 CTCCTCTGCACACCGTTGGATTTGGGCACTGCAGGCATGGAGCTGTGTGGACTTC 288
QY 223 LeuLeuAlaHisGlyAlaAspValAspSerLeuAspValIysAlaGlnThrProLeuPhe 242
DB 289 CTCATCCGGAAGGGCGCGAGTGGATCTGTGGACGTAAAGGACAGACGCGCCCTGTAT 348
QY 243 ThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSer 262
DB 349 GTGGCTGTGGTGAACGGGACCTAGAGAGTACCAGATCTCTCGAAGCTGGCGGGAC 408
QY 263 ProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAla 282
DB 409 CCCACGGAGCGCGACCATCCGACGCCCTCTTACCAGCCTCTCGCTGGGCGCG 468
QY 283 ValAlaIleLeuGlnLeuLeuAspHisGlyAlaGluAlaAsnValIysAlaLysLeu 302
DB 469 GCAGACATCTGAAGCCCTCATCAGGTACGGGCTGTGTTGACGTCAACACACCTTC 528
QY 303 -----ProValTrpAlaSerAsnIleAlaSer-----CysSerGly 314
DB 529 ACTCTGTATGTCAGCTCGATTCTCCGGCGGCTCACCTCTCTGTGTCTGCTGC----- 582
QY 315 ProLeuTyrLeuAlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHis 334
DB 583 CCTTGTATCATCAGCAGCCTACCAACCTCCAGTGTTCGGGTCTCTCTCTGGCT 642
QY 335 GlyAlaAspProAspTyrAsnCys-----ThrAspGlnGlyLeuLeuAlaArg 350
DB 643 GCGCGAACCTGACTCACTCACTGCAATGCTGTCTCAACACACAGGAGTCTACAGG--- 699
QY 351 ValProArgProArgThrLeuLeuGluIleCysLeuHisAsnCysGluProGluTyr 370
DB 700 ---GGCTCCCTCGGGCGCTCATGATGCTGTCTCGCCACCGCTGTGAGCAGCCTTC 756
QY 371 IleGlnLeuLeuIleAspPheGlyAlaAsnIleTyrLeu-----ProSerLeuSer 387
DB 757 GTGAGCTCTGTGTAGATTTGGAGCAACCTGATCTAGTGAAGTGGGAATCGCTGGCG 816
QY 388 LeuAspLeuThrSerGln-----AspAspLysGlyIleAlaLeuLeuGlnAla 404
DB 817 CCAGAGTCGAGGAGAGAGAAAGTGGACCTCGAGGCTTCGAGCTCTTAAAGAGGCC 876
QY 405 ArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgAlaLeuCys 424
DB 877 AGAAGTGTTCACAGAACCTTGCTGTCTGTGCGGCTGTGAGAGAGAGCTCTTTGGC 936
QY 425 GlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProMetLeuIleSerTyr 444
DB 937 AAA---CACCGGCTTCATCTGATTCCTCTGCTGCTCTGCGACCCCATAAAGAGTTT 993
QY 445 LeuLysHisGln 448
DB 994 CTACTCCATGAG 1005
RESULT 12
ABV29074
ID ABV29074 standard; cDNA; 1433 BP.
XX AC ABV29074;
XX AC ABV29074;
XX AC ABV29074;
DT 16-SEP-2002 (first entry)
XX DE Human prostate expression marker cDNA 29065.
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX KW pharmacogenomic marker; gene; ss.
XX OS Homo sapiens.
XX PN WO200160860-A2.
```

```
XX 23-AUG-2001.
XX 20-FEB-2001; 2001WO-US05171.
XX 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Schlegel R, Endege WO, Monahan JB;
XX WPI; 2001-662795/76.
DR Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX Claim 1; Page 6164-6165; 11750pp; English.
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX Sequence 1433 BP; 280 A; 377 C; 410 G; 336 T; 30 other;
SQ Alignment Scores:
Pred. No.: 3,96e-41 Length: 1433
Score: 499.00 Matches: 119
Percent Similarity: 57.89% Conservative: 57
Best Local Similarity: 39.14% Mismatches: 102
Query Match: 21.30% Indels: 26
DB: 23 Gaps: 9
US-09-941-831A-20 (1-449) x ABV29074 (1-1433)
```

```
QY 166 GluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeu 185
DB 109 GAGGACACGAGGCTCCATGATGAGCTTACGTCCGGGACCTCCAGACCTCCAGAGCCTA 168
QY 186 LeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGlyTrp-----Gly 202
DB 169 TTGCAAGAGGAGGAGCTTACCGGAGCCGCATCAACGAGAGAGTCTGTCTGGTGTGGCTGG 228
QY 203 ValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnVal 222
DB 229 CTCCTCTGCACACCGTTGGGAATTCGGGCATCGAGGCGCATGGGAGCTGTGTGGACTTC 288
QY 223 LeuLeuAlaHisGlyAlaAspValAspSerLeuAspValIysAlaGlnThrProLeuPhe 242
DB 289 CTCATCCGGAAGGGCGCGAGTGGATCTGTGGACGTAAAGGACAGACGCGCCCTGTAT 348
QY 243 ThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSer 262
DB 349 GTGGCTGTGGTGAACGGGACCTAGAGAGTACCAGATCTCTCGAAGCTGGCGGGAC 408
QY 263 ProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAla 282
```

```
Db 409 CCCACGGAGCGCGCACCATCGCAGCACCCCTGTCTACCCAGCCTCTCGCGTGGCGCG 468
Qy 283 ValAlaIleLeuGlnLeuLeuAAspHisGlyAlaGluAlaAsnValLysAlaLysLeu 302
Db 469 GCAGACATCTGAAGCGCCCTCATCAGGTACGGGGCTGATGTTACGCTCAACACCACTG 528
Qy 303 -----ProValTrpAlaSerAsnIleAlaSer-----CysSerGly 314
Db 529 ACTCCTGATGTCAGCCTCGATCTCCCGCGGCTCACCTCTCTGTGGTCTGC----- 582
Qy 315 ProLeuTyrlLeuAlaValTyrlGlyHisLeuAAspCysPheArgLeuLeuLeuHis 334
Db 583 CCCTTGATACATCAGCGCAGCTACCAACCTCAGTGTCTCCGGCTCTCTCTCTCTCTCT 642
Qy 335 GlyAlaAspProAspTyrlAsnCys-----ThrAspGlnGlyLeuLeuAlaArg 350
Db 643 GCGCGGAACCTGACTTCACTCAATGTCGTGTCAACACAGAGGATTTACAGG--- 699
Qy 351 ValProArgProArgThrLeuLeuGluLeuGluLeuHisAsnCysGluProGluTyrl 370
Db 700 ---GGCTCCCTGGGCGGTGATGATGCTGTCTGCGCCACGGCTGTGAGCGAGCCTTC 756
Qy 371 IleGlnLeuLeuLeuAspPheGlyAlaAsnIleTyrlLeu-----ProSerLeuSer 387
Db 757 GTGAGCCTGCTGTAGAAATTGAGGCAACCTGAATCTAGTGAAGTGGGAATCGCTGGGC 816
Qy 388 LeuAspLeuThrSerGln-----AspAspLysGlyIleAlaLeuLeuGlnAla 404
Db 817 CCNAGTCGAGGAGAGAGAAAGTGGACCTTGAGGCTTGCGAGTCTTTAAAGAGGCC 876
Qy 405 ArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgAlaLeuCys 424
Db 877 AGAAGTGTCCAGAACCTTGTCTGTCTGCGGTGTGGCTGTGAGAGAGCTCTTGGC 936
Qy 425 GlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMetLeuLeuSerTyrl 444
Db 937 AAA---CACCGGCTTCATCTGATTCCTCTCGCTCTGCTCTGCGAGACCCCATAAAGATT 993
Qy 445 LeuLysHisGln 448
Db 994 CTACTCCATGAG 1005

RESULT 13
AAD49617
ID AAD49617 standard; cDNA; 1509 BP.
XX AC AAD49617;
XX DT 24-MAR-2003 (first entry)
XX DE Human cytoskeleton-associated protein, CSAP-28 cDNA.
XX KW Human; cytoskeleton-associated protein; CSAP-28; atherosclerosis;
XX KW cancer; gene therapy; gene; ss.
XX OS Homo sapiens.
PH Key Location/Qualifiers
FT CDS 12..1025
FT /*tag= a
FT /product= "Human CSAP-28"
XX PN WO200279404-A2.
XX PD 10-OCT-2002.
XX PF 25-MAR-2002; 2002WO-US09288.
XX PR 29-MAR-2001; 2001US-280508P.
XX PR 03-APR-2001; 2001US-281323P.
XX PR 13-APR-2001; 2001US-283769P.
XX PR 04-MAY-2001; 2001US-288609P.
```

```
PR 10-MAY-2001; 2001US-290518P.
PR 18-MAY-2001; 2001US-291870P.
PR 29-MAY-2001; 2001US-294451P.
XX (INCY-) INCYTE GENOMICS INC.
XX Hafalia AJA, Tang TY, Yue H, Khan FA, Ison CH, Baughn MR;
PI Warren BA, Duggan BM, Thangavelu K, Honchell CD, Azimzai Y;
PI Elliott VS, Burford N, Ding L, Yue H, Becha S, Emerling BM;
PI Richardson TW, Lee SY, Bandman O, Lal PG, Lee S, Gietzen KJ;
PI Wallia NK, Griffin JA, Lee EA, Swarnakar A, Ring HZ, Jones KA;
XX WPI; 2003-092894/08.
DR P-PSDB; AAE32130.
XX New human cytoskeleton-associated proteins, useful for preparing a
PT composition for diagnosing or treating a disease or condition
PT associated with decreased expression or overexpression of functional
PT CSAP e.g., cancer
XX Claim 5; Page 232-233; 233pp; English.
XX The invention relates to new human cytoskeleton-associated protein
CC (CSAP) and its polynucleotide. The polypeptide is useful for preparing
CC a composition for diagnosing or treating a disease or condition
CC associated with decreased expression or overexpression of functional
CC CSAP e.g. atherosclerosis or cancer. The present sequence is human
CC CSAP-28 cDNA. The invention is useful in gene therapy.
XX Sequence 1509 BP; 340 A; 396 C; 449 G; 324 T; 0 other;
```

```
Alignment Scores:
Pred. No.: 4,27e-41 Length: 1509
Score: 499.00 Matches: 119
Percent Similarity: 57.89% Conservative: 57
Best Local Similarity: 39.14% Mismatches: 102
Query Match: 21.30% Indels: 26
DB: 25 Gaps: 9
```

```
US-09-941-831A-20 (1-449) x AAD49617 (1-1509)
Qy 166 GluLysGlnAlaLeuAsnGlnAlaValTyrlAspAsnAspSerTyrlThrLeuAspGlnLeu 185
Db 126 GAGGACACGAGGCTCCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 185
Qy 186 LeuArgGlnGluArgTyrlLysArgPheIleAsnSerArgSerGlyTrp-----Gly 202
Db 186 TTGCAAGAGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 245
Qy 203 ValProGlyThrProLeuArgLeuAlaAlaSerTyrlGlyHisLeuSerCysLeuGlnVal 222
Db 246 CTCCTCTGCACACCTTGGGAATCGGCCACTCGGCCACTCGGCCACTCGGCCACTCGGCC 305
Qy 223 LeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPhe 242
Db 306 CTCATCCGGAAGGGGCGAGGTGATCTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 365
Qy 243 ThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGlnAlaGlyAlaSer 262
Db 366 GTGGCTGTGGTGAACGGGACCTAGAGAGATCCCAAGATCTCTCGAAGCTCGCGCGGAC 425
Qy 263 ProGlyGlySerIleTyrlAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAla 282
Db 426 CCCAAGCGAAGCGGCGGACCATCGAGCACCCTCTGTACCACGCTCTCGCGTGGCGCGG 485
Qy 283 ValAlaIleLeuGlnLeuLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeu 302
Db 486 GCAGACATCTCTGAAGCCCTCATCAGGTACGGGGCTGATGTTGACGTCAACCAACACCTG 545
Qy 303 -----ProValTrpAlaSerAsnIleAlaSer-----CysSerGly 314
Db 546 ACTCTGTATGTCACGCTCGATTTCTCCCGGGGCTTCTCTCTCTCTCTCTCTCTCTCTCT 599
```



```

Db 555 ACTCTGATGTCAGCTCCCGCGGCTCACCTCCTTGGTGGTCTGC----- 608
Qy 315 ProLeuTyrLeuAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHis 334
Db 609 CCCTTGATACATCAGCGAGCTACCAACCTCCAGTGTTCGGGCTCTCTCTGGCT 668
Qy 335 GlyAlaAspProAspTyrAsnCys-----ThrAspGlnGlyLeuLeuAlaArg 350
Db 669 GCGCGAACCCTGACTCACTGATGCTGCTCAACACAGGATTCACAGG--- 725
Qy 351 ValProArgProArgThrLeuLeuGluLeuHisCysLeuHisAsnCysGluProGluTyr 370
Db 726 ---GGCTCCCTGGTGGTGCATGATGCTGTTCTGCGCACCGCTGTGAGCGAGCTTC 782
Qy 371 IleGlnLeuLeuLeuAspPheGlyAlaAsnIleTyrLeu-----ProSerLeuSer 387
Db 783 GTGAGCTGTGCTAGTAATTTGAGGCAACCTCAATCTAGTGAAGTGGGAATCGCTGGC 842
Qy 388 LeuAspLeuThrSerGln-----AspAspGlyGlyIleAlaLeuLeuGlnAla 404
Db 843 CCAGATCGAGGAGAGAAAGTGGACCTTGAGCGCTTGACGCTCTTTAAAGAGGCC 902
Qy 405 ArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgAlaLeuCys 424
Db 903 AGAAGTGTTCACAGAACCTTGTCTGTCTGTCGCGTGTGCTGTGAGAGAGCTCTTGGC 962
Qy 425 GlnAlaGlyClnProGlnAlaIleAsnGlnLeuAspIleProProMetLeuIleSerTyr 444
Db 963 AAA---CACCGGCTTCATCTGATCTCTGCTGCTGCTGCGACCCCATAAAGAAGTTT 1019
Qy 445 LeuLysHisGln 448
Db 1020 CTACTCATGAG 1031

RESULT 15
AAV38672
ID AAV38672 standard; DNA; 2019 BP.
XX
AC AAV38672;
XX
DT 27-OCT-1998 (first entry)
XX
DE Mus musculus SOCS7 gene.
XX
KW SOCS; suppressor of cytokine signalling; PCR primer;
KW autoimmune disease; diagnosis; cancer; treatment;
KW cytokine mediated cellular responsiveness; hyperimmunity;
KW immunosuppression; allergies; hypertension; ss.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 2..1054
FT FT /*tag= a
FT FT /product= SOCS7 protein
XX
PN WO9820023-A1.
XX
PD 14-MAY-1998.
XX
PF 31-OCT-1997; 97WO-AU00729.
XX
PR 14-FEB-1997; 97AU-0005117.
XX
PR 01-NOV-1996; 96AU-0003384.
XX
PR (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
XX
PI Alexander WS, Hilton DJ, Metcalf D, Nicholson SE;
PI Nicola NA, Richardson RT, Starr R, Viney EM, Willson TA;
XX
DR WPI; 1998-286854/25.
DR P-PSDB; AAW62621.

```

```

XX
PT Suppressor of cytokine signalling proteins - useful to treat
PT disease, injury or abnormality involving cytokine mediated cellular
PT responsiveness e.g. hyperimmunity, immunosuppression, allergies and
PT hypertension
XX
PS Claim 14; Page 143-144; 325pp; English.
XX
CC The sequence is that of a gene encoding a suppressor of cytokine
CC signalling protein (SOCS). SOCS can be used to screen for naturally
CC occurring antibodies to SOCS, which may occur, e.g. in some autoimmune
CC diseases. Alternatively, specific antibodies can be used to
CC screen for SOCS, which is useful as a knowledge of SOCS levels
CC may be important for the diagnosis of certain cancers. Soluble
CC SOCS polypeptides can be used to treat disease, injury or
CC abnormality involving cytokine mediated cellular responsiveness,
CC e.g. hyperimmunity, immunosuppression, allergies and hypertension.
XX
SQ Sequence 2019 BP; 427 A; 510 C; 545 G; 528 T; 9 other;

Alignment Scores:
Pred. No.: 5,85e-39 Length: 2019
Score: 480.00 Matches: 114
Percent Similarity: 57.72% Conservative: 58
Best Local Similarity: 38.26% Mismatches: 104
Query Match: 20,49% Indels: 22
DB: 19 Gaps: 8

US-09-941-831A-20 (1-449) x AAV38672 (1-2019)
Qy 170 LeuAsnGlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeuLeuArgGlnGlu 189
Db 167 CTCATGATGACGCTATGTAGGGACCTCCAGACCTCAGAACCTACTGCAAGAGGAG 226
Qy 190 ArgTyrLysArgPheIleAsnSerArgSerGlyTyr-----GlyValProGlyThr 206
Db 227 AGCTACCGGAGCGCATCAATGAGAAGTCTGTGTGTGCTGCGGTGGCTTCCTGCACA 286
Qy 207 ProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAlaHis 226
Db 287 CCACCTGAGATCGCAGCCACTCGAGGCCATGGGAAGTGTGTGACTTCTCATACGCAA 346
Qy 227 GlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPheThrAlaValSer 246
Db 347 GGGGCGAGGTGACCTGTGTGATGTCAAGGGGAGAGCTCCCTGTATGTGCTGTAGTG 406
Qy 247 HisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerProGlyGlySer 266
Db 407 AACGGGCACTTGGAGAGCAGCTGAGATCCTTTTGGAAAGCTGTGTGATCCCAACGGCAG 466
Qy 267 IleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaValAlaLeu 286
Db 467 CGGCACCCAGCAGCAGCTCTGTGTACCATGCTTTCGTGTGGTAGGAGCAGCATCTCG 526
Qy 287 GlnGluLeuLeuHisGlyAlaGluAlaAsnValLysAlaLysLeu----- 302
Db 527 AAGGCTCTTATCAGGTATGGGACAGATGTTGATGTCAACCATCATCTGATTTCTGACACC 586
Qy 303 ---ProValTyrAlaSerAsnIleAlaSer-----CysSerGlyProLeuTyrLeu 318
Db 587 CGGCCCCCTTTTTCACGGCGCTAACCTCCTTGGTGGTCTGT-----CCTCTATACATC 640
Qy 319 AlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspPro 338
Db 641 AGTGTGCTCTACCAATACCTTCAGTGTCTCAGGTGCTCTTTCAGGCTGGGCAATCTCT 700
Qy 339 AspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgValProArg-----ProArgThr 356
Db 701 GACTTCAATTGCAATGGCCCTCTCAACACCCAGGAGTCTACAGGGGATCCCTCGGTGT 760
Qy 357 LeuLeuGluIleCysLeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuLeuAsp 376
Db 761 GTCATGGATGCTCTCTGCGCCATGGCTGTGAAGCAGAGCCCTTCGTGAGTCTGTTGGTAGAG 820

```

```

QY 377 PheGlyAlaAsnIleTyrLeu-----ProSerLeuSerLeuAspLeuThrSerGln 393
Db 821 TTGGAGCCCAACCTGAACCTGGTGAAGTGGGAATCCCTGGGCCAGAGGCAAGAGCCAGA 880
QY 394 -----AspAspLysGlyIleAlaLeuLeuLeuGlnAlaArgAlaThrProArgSer 410
Db 881 AGAAGATGGATCCTGAGGCCCTTGCAGGTCTTTAAAGAGGCCCAAGATATTCCAGGACC 940
QY 411 LeuLeuSerGlnValArgLeuValValArgAlaLeuCysGlnAlaGlyGlnProGln 430
Db 941 TTGCTGAGTTGTGCCGGGTGGCTGTGAGAGAGACTCTTGGCAATATAC---CGACTGCAT 997
QY 431 AlaIleAsnGlnLeuAspIleProProMetLeuIleSerTyrLeuLysHisGln 448
Db 998 CTGGTCCCTCGCTGCCGCTGCCAGACCCCAAGAGTTTGTGCTTTATGAG 1051

```

Search completed: December 1, 2003, 08:00:21
Job time : 377 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 1, 2003, 07:53:58 ; Search time 2294 Seconds
(without alignments)
644.863 Million cell updates/sec

Title: US-09-941-831A-20
Perfect score: 2343
Sequence: 1 MRESGWKLIDPDRMGIM.....QAINQLDIPPLMLSYLKHQL 449

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2190069 seqs, 1647345023 residues

Total number of hits satisfying chosen parameters: 4380138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlpl

-Q=/cg2_1/USPTO.spool_p/US0941831/runat_01122003_073528_7884/app_query.fasta_1.647
-DB=Published Applications NA -QMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blom62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US0941831_@CGN_1_1_95 @runat_01122003_073528_7884
-NCPUS=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FCAPOP=6 -FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

- 1: /cg2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cg2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cg2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cg2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cg2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cg2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cg2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cg2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cg2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cg2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cg2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cg2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cg2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 14: /cg2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 15: /cg2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 16: /cg2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 17: /cg2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2343	100.0	1878	10	US-09-941-831-6 Sequence 6, Appli

2	2343	100.0	1878	12	US-10-207-175-13
3	1690	72.1	1222	10	US-09-908-193-13
4	1629	69.5	1069	10	US-09-908-193-11
5	606.5	25.9	522	10	US-09-833-381-62
6	480	20.5	2019	10	US-09-908-805B-24
7	265	11.3	1118	12	US-10-314-321A-60
8	265	11.3	1251	12	US-10-314-321A-64
9	264	11.3	3482	11	US-09-919-039-133
10	261.5	11.2	1206	14	US-10-037-270-34
11	246	10.5	990	9	US-09-884-870-3
12	246	10.5	1538	9	US-09-884-870-1
13	244.5	10.4	401	9	US-09-922-217-28
14	244.5	10.4	401	9	US-09-833-263-28
15	244.5	10.4	401	13	US-10-025-380-28
16	233.5	10.0	389	10	US-09-867-701-5266
17	233.5	10.0	419	10	US-09-908-805B-26
18	230	9.8	549	12	US-10-314-321A-58
19	230	9.8	616	12	US-10-314-321A-59
20	223	9.5	5482	12	US-10-199-937-138
21	217.5	9.3	4275	11	US-09-972-115A-5
22	217.5	9.3	4406	12	US-10-199-937-106
23	217.5	9.3	4992	12	US-10-199-937-134
24	217.5	9.3	5002	12	US-10-199-937-132
25	217.5	9.3	6018	12	US-09-849-602-11
26	216	9.2	6096	12	US-10-133-013-169
27	215	9.2	1568	12	US-10-199-937-98
28	215	9.2	3400	9	US-09-509-196A-1
29	215	9.2	3508	12	US-10-199-937-1
30	215	9.2	4127	12	US-10-199-937-100
31	215	9.2	5778	14	US-10-163-587A-14
32	214	9.1	2971	12	US-10-199-937-92
33	214	9.1	3353	12	US-10-199-937-88
34	214	9.1	3799	12	US-10-199-937-90
35	213.5	9.1	4491	9	US-09-841-835-7
36	209	8.9	4657	9	US-09-841-835-9
37	208.5	8.9	3984	12	US-10-199-937-3
38	208.5	8.9	4134	9	US-09-841-835-1
39	208.5	8.9	4134	11	US-09-972-115A-7
40	207	8.8	1712	14	US-10-037-270-1101
41	205	8.7	5175	10	US-09-964-899-42
42	205	8.7	14770	12	US-09-873-367C-230
43	203	8.7	6207	12	US-10-205-194-118
44	200	8.5	1066	9	US-09-835-788A-8
45	200	8.5	1066	12	US-10-175-042-8

ALIGNMENTS

RESULT 1

US-09-941-831-6
; Sequence 6, Application US/09941831
; Patent No. US20020160493A1
; GENERAL INFORMATION:
; APPLICANT: Ebner et al.
; TITLE OF INVENTION: P704991
; FILE REFERENCE: Serine/Threonine Phosphatase Polynucleotides, Polypeptides, and Antib
; CURRENT APPLICATION NUMBER: US/09/941,831
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: PCT/US01/06256
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/186,350
; PRIOR FILING DATE: 2000-03-02
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1878
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-941-831-6

Alignment Scores:	1.54e-298	Length:	1878
Pred. No.:	2343.00	Matches:	449

QY 61 PheArgSerLysGluArgValProValLeuSerTyrLeuTyrLysGluAsnAlaAla 80
DB 547 TTCAGAGTAAGAAGCGTCTCTGTCTCTCTACCTCTACAAAGAGACAACTGCTGCC 606
QY 81 IleCysArgCysSerGlnProLeuSerGlyPheTyrThrArgCysValAspAspGluLeu 100
DB 607 ATTTGCGGCTGAGCAGCGCTCTCTGTGATTTTACACTCGCTGTGTAGATGATGAGCTC 666
QY 101 LeuLeuGluAlaIleSerGlnThrAsnProGlySerGlnPheMetTyrValValAspThr 120
DB 667 TTGTTGGAGGCATTAGCCAAACAAACCAGGAGCCAGTTTATGTATGTGTGACACA 726
QY 121 ArgProLysIleTrpHisPheLeuValLeuIleMetArgIleValLeuGlnLeuAlaLys 140
DB 727 AGACCAAGATCTGGCATTTCTTGTGCTCATATGAGATAGTTCTCCAATTACCCAA 786
QY 141 MetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGlu 160
DB 787 ATGAACCTCATGGACATCAACCAAGATCTTCTCCCTCTGACGCCGACAAAGGAGGAG 846
QY 161 AspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyr 180
DB 847 GACACTGACACAGAGAGAGAGCGCTCTCAATCAAGCAGTGTATGACAAACACTCCTAT 906
QY 181 ThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGly 200
DB 907 ACTTTGGACCACTTTTGGCCAGAGCGTTCACAAACGTTTCATCAACGACAGAGTGGC 966
QY 201 TrpGlyValProGlyThrProLeuArgLeuAlaIleSerTyrGlyHisLeuSerCysLeu 220
DB 967 TGGGGTGTCTCTGGACACCTTGGCTTGGCTTCTTATGGCCACTTGAAGTGTGTG 1026
QY 221 GlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrPro 240
DB 1027 CAAGTCCTCTTAGCCCATGGTCTGATGTGTGACGCTTGGATGTCAAGGCACAGACGCA 1086
QY 241 LeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGly 260
DB 1087 CTTTTCACGTCTGTCATGTCATGCCCATCTGGACTGTGTACGTGTCTTTTGAAGCTGT 1146
QY 261 AlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAsp 280
DB 1147 GCCTCTCTCTGGTGTAGCATCTACAACACTGTCTCCGCTGCTCACAGCTGCCGTGAT 1206
QY 281 GlyAlaValAlaIleLeuGlnLeuLeuAspHisGlyAlaGluAlaAsnValLysAla 300
DB 1207 GGTGCTGTTGTCTATCCTGCAGGAGCTCTTAGACCATGGTGTGCAGAGGCCAACGCTCAA 1266
QY 301 LysLeuProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAla 320
DB 1267 AAACCTACAGTCTGGGCATCAACATAGCTTATGTCTGGCCCCCTCTATTGTGGCCGA 1326
QY 321 ValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyr 340
DB 1327 GTCTAGGGGACCTGGACTGTTTCCGCTGCTTTTGTCTCCAGGGGACACCTGACTAC 1386
QY 341 AsnCysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIle 360
DB 1387 AACTGACCTGACAGCGCCCTATTGGCTGCTGTCCTCCAAAGACCCCGCACCTCTCTGAAATC 1446
QY 361 CysLeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuIleAspPheGlyAlaAsn 380
DB 1447 TGCCTCCATCAATATGTGAGCCAGATATATCCAGCTGTAACTGTTTGGTGTCTAAT 1506
QY 381 IleTyrLeuProSerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeu 400
DB 1507 ATCTACCTTCCATCTCTCTCCCTTGACCTGACCTCACAGATGATAAAGCATTCGATTG 1566
QY 401 LeuLeuGlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArg 420
DB 1567 CTGCTACAGGCCCGAGCCACTCCACGGTCACCTTCTATCATCAGGTCGGTTTAGTCGTCGC 1626

QY 421 ArgAlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMet 440
DB 1627 AGAGCTTGTGCCAGGCTGGCCAGCCACCAAGCCATCAACCAAGCTGGATATTCTCTCCCATG 1686
QY 441 LeuIleSerTyrLeuLysHisGlnLeu 449
DB 1687 TTGATTAGCTACCTAAAACACCAACTG 1713
RESULT 3
US-09-908-193-13
; Sequence 13, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHIMKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
; APPLICANT: PADIGARU, MURALIDHARA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 21402-062
; CURRENT APPLICATION NUMBER: US/09/908,193
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: 60/220,273
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/221,650
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/221,233
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/220,912
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/218,875
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,870
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,901
; PRIOR FILING DATE: 2000-07-18
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1222
; TYPE: DNA
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: NOVX nucleic acid
; OTHER INFORMATION: with homology to ankyrin repeat-containing-
; OTHER INFORMATION: (ASB-1) like protein
US-09-908-193-13
Alignment Scores:
Pred. No.: 1,55e-212 Length: 1222
Score: 1690.00 Matches: 326
Percent Similarity: 99.70% Conservative: 1
Best Local Similarity: 99.99% Mismatches: 1
Query Match: 72.13% Indels: 0
DB: 10 Gaps: 0
US-09-941-831a-20 (1-449) x US-09-908-193-13 (1-1222)
QY 122 ProLysIleTrpHisPheLeuValLeuIleMetArgIleValLeuGlnLeuAlaLysMet 141
DB 102 CCACAGATCTGGCATTTCTTGTGCTCATATGAGATAGTTCTCCAATTAGCCAAAGATG 161
QY 142 AsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGluAsp 161
DB 162 AACCTCATGGACATCACCAAGATCTTCTCCCTCTGACGCCGACCAAGAGGAGGAGGAC 221
QY 162 ThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThr 181
DB 222 ACTGACACAGAGAGAGAGCGCTCTCAATCAAGCAGTGTATGACACGACTCTCTATAT 281
QY 182 LeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSerGlyTyr 201
DB 282 TTGGACCAAGCTTTTGGCCAGGAGGCTTACAAACGTTTTCATCAACAGCAGGAGTGGCTGG 341

QY 202 GlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGln 221
 Db 342 GGTGTTCTCGGACACCCCTGGCTGGCTCTTATGGCCACTTGAGCTGTTGGCAA 401
 QY 222 ValLeuAlaHisGlyAlaAspValAspSerLeuAspValysAlaGlnThrProLeu 241
 Db 402 GTCTCTTAGCCCATGGTCTGATGTTGACAGCTTGATGTCAAGGCACAGCCACTT 461
 QY 242 PheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAla 261
 Db 462 TTCACTGCTGTGAGTCAATGCCATCTGGACTGTGTACGTGTCTTTTGAAGCTGGTGC 521
 QY 262 SerProGlyGlySerIleTyrHisLeuSerProValLeuThrAlaAlaArgAspGly 281
 Db 522 TCTCTGGGTGGAGCATACAACTGTCTCCGCTGCTCACAGCTCCCGGTGATGT 581
 QY 282 AlaValAlaIleLeuGlnLeuLeuAspHisGlyAlaGluAlaAsnValysAlaLys 301
 Db 582 GCTGTTGCTATCTCGAGAGCTCTTAGACCATGGTGCAGGCCAACGTCAGCTTAA 641
 QY 302 LeuProValTrpAlaSerHisLeuAlaSerCysSerGlyProLeuTyrLeuAlaVal 321
 Db 642 CTACCACTCTGGGCATCAAACTAGCTTCATGTTCTGGCCCTCTATTTGGCGCAGTC 701
 QY 322 TyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsn 341
 Db 702 TAGGGGCCTGGACTGTTTCCGCTGCTTTTGTCTCCAGGGGCAGACCTGACTACAAC 761
 QY 342 CysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluLeuCys 361
 Db 762 TGCACGTACAGGGCCCTATTTGGCTGCTGTCCTCCAGACCCCGCACCTCTTGAATCTGC 821
 QY 362 LeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuLeuAspPheGlyAlaAsnIle 381
 Db 822 CTCCATCAATATGTGAGCCAGAGTATATCCAGCTGTTAATCGATTTTGGTGTCAATATC 881
 QY 382 TyrLeuProSerLeuSerLeuAspLeuThrSerGlnAspAspGlyIleAlaLeuLeu 401
 Db 882 TACCTTCCATCTCTCTCCCTGACCTGACCTCACAGATGATAAAGGCCATTCATTGCTG 941
 QY 402 LeuGlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgArg 421
 Db 942 CTACAGCCCGGAGCCACTCCAGGTCACTTCTATCAAGCTCCGTTAGTCTGCCGAGA 1001
 QY 422 AlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProMetLeu 441
 Db 1002 GCCTTGTGCAGCTGCCAGCCACCAAGCCATCAACAGCTGGATATCTCCCATGTTG 1061
 QY 442 IleSerTyrLeuLysHisGlnLeu 449
 Db 1062 ATTAGCTACCTAAACACCAACTG 1085

RESULT 4

US-09-908-193-11
 ; Sequence 11, Application US/0908193
 ; Publication No. US20020192748A1
 ; GENERAL INFORMATION:
 ; APPLICANT: RASTELLI, LUCA
 ; APPLICANT: SHIMKETS, RICHARD A.
 ; APPLICANT: ZERHUSEN, BRYAN
 ; APPLICANT: MALYANKAR, URIEL M.
 ; APPLICANT: PADIGARU, MURALIDHARA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
 ; FILE REFERENCE: 21402-062
 ; CURRENT APPLICATION NUMBER: US/09/908,193
 ; CURRENT FILING DATE: 2001-07-18
 ; PRIOR APPLICATION NUMBER: 60/220,273
 ; PRIOR FILING DATE: 2000-07-24
 ; PRIOR APPLICATION NUMBER: 60/221,650
 ; PRIOR FILING DATE: 2000-07-28
 ; PRIOR APPLICATION NUMBER: 60/221,233
 ; PRIOR FILING DATE: 2000-07-27

; PRIOR APPLICATION NUMBER: 60/220,912
 ; PRIOR FILING DATE: 2000-07-26
 ; PRIOR APPLICATION NUMBER: 60/218,875
 ; PRIOR FILING DATE: 2000-07-18
 ; PRIOR APPLICATION NUMBER: 60/218,870
 ; PRIOR FILING DATE: 2000-07-18
 ; PRIOR APPLICATION NUMBER: 60/218,901
 ; PRIOR FILING DATE: 2000-07-18
 ; NUMBER OF SEQ ID NOS: 74
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 11
 ; LENGTH: 1069
 ; TYPE: DNA
 ; ORGANISM: Unknown Organism
 ; FEATURE:
 ; OTHER INFORMATION: Description of Unknown Organism: NOVX nucleic acid
 ; OTHER INFORMATION: with homology to ankyrin repeat-containing-
 ; OTHER INFORMATION: (ASB-1) like protein
 ; US-09-908-193-11

Alignment Scores:

Pred. No.: 1.43e-204 Length: 1069
 Score: 1629.00 Matches: 318
 Percent Similarity: 97.55% Conservative: 0
 Best Local Similarity: 97.55% Mismatches: 0
 Query Match: 69.53% Indels: 8
 DB: 10 Gaps: 1

US-09-941-831A-20 (1-449) x US-09-908-193-11 (1-1069)

QY 112 MetArgIleValLeuGlnLeuAlaLysMetAsnLeuMetAspIleThrIlePheSer 151
 Db 81 ATGAAATAGTGTCTCCAATTAGCCAAAGATGAACCTCATGGACATCACCAGATCTTCTCC 140
 QY 152 LeuLeuGlnProAspLysGluGluAspThrAspThrGluGluGlnAlaLeuAsn 171
 Db 141 CTCCTGCAGCCGCAAGGAGGAGGAGGACACTGACACAGAGGAGAGCAGCTCTCAAT 200
 QY 172 GlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeuLeuArgGlnIleArgTyr 191
 Db 201 CAAGCAGTGTATGACACAGCACTCTCTATATTTGGACCAAGCTTTTGGCCAGAGCGTTAC 260
 QY 192 LysArgPheIleAsnSerArgSerGlyTyrGlyValProGlyThrProLeuArgLeuAla 211
 Db 261 AAACGTTTTCATCAACAGCAGAGAGTGGCTGGGTGTTCTGGACACCCCTTGGCTGGCT 320
 QY 212 AlaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAlaHisGlyAlaAspValAsp 231
 Db 321 GCTTCTTATGGCCACTTGAGCTGTTTGAAGTCTCTTAGCCCATGGTGTGATGTTGAC 380
 QY 232 SerLeuAspValLysAlaGlnThrProLeuPheThrAlaValSerHisGlyHisLeuAsp 251
 Db 381 AGCTTGGATGTCAAGGCACAGACGCCACTTTTCACTGTCTGTCACTCATGGCCATCTGGAC 440
 QY 252 CysValArgValLeuLeuGluAlaGlyAlaSerProGlyGlySerIleTyrHisLeuAsnCys 271
 Db 441 TGTGTACGTGTCTTTTGAAGCTGGTGCCTCTCTCTGGTGGTAGCATCTTACAACTGT 500
 QY 272 SerProValLeuThrAlaAlaArgAspGlyAlaValAlaIleLeuGlnLeuLeuAsp 291
 Db 501 TCTCCGTGTCTCACAGCTGCCGTGATGGTGTCTGTGTGTATCTCTCAGAGAGCTCTAGAC 560
 QY 292 HisGlyAlaGluAlaAsnValLysAlaLysLeuProValTrpAlaSerHisLeuAlaSer 311
 Db 561 CATGTGCAGAGGCCCAAGCTCAAAAGCTAACTACCACTGTGGGCATCAAACTAGCTTCA 620
 QY 312 CysSerGlyProLeuTyrLeuAlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeu 331
 Db 621 TGTTCGCCCCCTCTATTTGGCCGAGTCTACGGGACCTTGGACTGTTCGCTGCTT 680
 QY 332 LeuLeuHisGlyAlaAspProAspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgVal 351
 Db 681 TTGCTCCAGGGGAGACCTTGACTACAACATGCACTGACAGGAGGCTATTGCTGCTGTC 740


```
Db 287 CCCTAGGATCGCAGCCACTGCGGCACTGGAACCTGTGTGACCTTCCTCATACGCAAA 346
Qy 227 GlyAlaAspValAspSerLeuAspValValysAlaGlnThrProLeuPheThrAlaValSer 246
Db 347 GGGCCGAGGTGACCTGGTGGATGCAAGGGGAGAGTCCCTGTATGTGGTGTAGTG 406
Qy 247 HisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerProGlySer 266
Db 407 AACGGGCACCTTGGAGCAGCTGAGATCCTTTGGAGCTGGTGTGATCCCAACGGCAGC 466
Qy 267 IleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaValAlaLeu 286
Db 467 CGCACACACCGCAGCACTCTGCTGTACCATGCTCTGTGGTAGGAGCAGCATCCTG 526
Qy 287 GlnGluLeuLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeu 302
Db 527 AAGGCTCTTATCAGGTATGGGAGATGTTGATGTCACCATCATCTGAATTCGTGACAC 586
Qy 303 ---ProValTrpAlaSerAsnIleAlaSer-----CysSerGlyProLeuTyrLeu 318
Db 587 CGGCCCCCTTTTACGGCGGCTAACCTCTCTGTGTGTGTGT-----CCTCTATACATC 640
Qy 319 AlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuLeuHisGlyAlaAspPro 338
Db 641 AGTGTCTCTTACCATACCTTCACTGTCTTGTGAGGCTGTCTTGTGAGGCTGGGCAATCCT 700
Qy 339 AspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgValProArg-----ProArgThr 356
Db 701 GACTTCATTTGCAATGGCCCTGTCAACACCCAGGAGTTCTACAGGGATCCCTCGGTGT 760
Qy 357 LeuLeuGluLeuLeuCysLeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuLeuAsp 376
Db 761 GTCATGATGCTCTCTCGCCCATGGCTGTGTAAGCAGCCTTCGTGAGTCTGTTGGTAGAG 820
Qy 377 PheGlyAlaAsnIleTyrLeu-----ProSerLeuSerLeuAspLeuThrSerGln 393
Db 821 TTTGGAGCCCAACCTGAACCTGTGTGAAGTGGGAATCCTCTGGCCGACAGGAGCAGAGA 880
Qy 394 -----AspAspLysGlyIleAlaLeuLeuLeuGlnAlaAlaArgAlaThrProArgSer 410
Db 881 AGAAAGATGATCTCGAGGCTTGCAGGCTTTAAAGAGCCAGAGATATTCACGAGCC 940
Qy 411 LeuLeuSerGlnValArgLeuValValArgAlaLeuCysGlnAlaGlyGlnProGln 430
Db 941 TTGCTCAGTTTGTGCGGGTGGCTGTGAGAAGAGCTCTTGGCAATAC---CGACTGCAT 997
Qy 431 AlaIleAsnGlnLeuAspLeuProProMetLeuLeuSerTyrLeuLysHisGln 448
Db 998 CTGGTTCCCTCGCTGCGCTGCCAGACCCCAATAAAGAGATTTTGTCTTTATGAG 1051
```

RESULT 7

```
US-10-314-321A-60
; Sequence 60, Application US/10314321A
; Publication No. US20030190648A1
; GENERAL INFORMATION:
; APPLICANT: Hitachi, Ltd.
; TITLE OF INVENTION: Gene Predicting Method
; FILE REFERENCE: 31010118U51
; CURRENT APPLICATION NUMBER: US/10/314,321A
; PRIOR FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: JP 2002-103333
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 60
; LENGTH: 1118
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: partial sequence of AL365356
US-10-314-321A-60
```

Alignment Scores:

```
Pred. No.: 3.13e-24 Length: 1118
Score: 265.00 Matches: 85
Percent Similarity: 47.15% Conservative: 39
Best Local Similarity: 32.32% Mismatches: 96
Query Match: 11.31% Indels: 44
DB: 12 Gaps: 10
```

US-09-941-831A-20 (1-449) x US-10-314-321A-60 (1-1118)

```
Qy 197 SerArgSerGlyTyrProGlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHis 216
Db 32 ACAGGTACCCCTGG---ACTCCATC-ACGCCCTTGCACGACCCAGTGTGACGGCCAG 87
Qy 217 LeuSerCysLeuGlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLys 236
Db 88 GCGCGGTGTGTGAGCTGTGTGTGGCGGTGGGCGCCAGTGTGATGCTCGCAACATCGAC 147
Qy 237 AlaGlnThrProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeu 256
Db 148 GGCAGCACCCCTCTCGCATGCTCGCTCGGCGCAGCATCGAGTGTGTGAAGCTCTTG 207
Qy 257 LeuGluAlaGlyAlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThr 276
Db 208 CTGTCTCTACGGGCGCAAGGTCAACCTCCCTCTGTAC---ACAGCGTCCCCCTGCACGAG 264
Qy 277 AlaAlaArgAspGlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAla 296
Db 265 GCTGTCATGAGCGGAGTTCGGAATGTGTGAGGCTCTTATTGACGTCGGG---GCC 318
Qy 297 AsnValLysAlaLysLeuProValTrpAlaSerAsnIleAlaSerCys-----SerGly 314
Db 319 AATCTGGAAGCGCAC-----GATTGCCATTTTGGGACC 351
Qy 315 ProLeuTyrLeuAlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHis 334
Db 352 CCTCTGCAGTGTGCTGTGCCCGGAGCATCTGGACTGTGTCAAAGTGTCTCAATGCA 411
Qy 335 GlyAlaAspProAspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgValProArgPro 354
Db 412 GGGGCCAAGCTGAAT-----GCG 429
Qy 355 ArgThrLeuLeuGluIleCysLeuHisHisAsnCysGlu-----ProGluTyrIle 371
Db 430 GCAAAAGCTTATGAGACTGCCCCCTTCAACACGCGGCCAAGGTCAAGAAATGTTGACCTCATC 489
Qy 372 GlnLeuLeuLeuAspPheGlyAlaAsnIleTyrLeu-----Pro 384
Db 490 GAGATGCTTATCGAGTTTGGCGGCAACATCTACGCCCGGACACCGCGGGAGAGCGCG 549
Qy 385 SerLeuSerLeuAspLeuThrSerGlnAspLysGlyIleAlaLeuLeuGlnAla 404
Db 550 TCT-----GACTACAGTGGAGCAGCAGCGCTCCGCCCAAGTGTTCGAGTACTAC 600
Qy 405 ArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgAlaLeuCys 424
Db 601 GAAAAGACACCTCTGACTCTGTACAGCTCTGTCCAGGTGAACTTGAAGGAGCCACTGCG 660
Qy 425 GlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMetLeuLeuSerTyr 444
Db 661 GTCCGAGGGGCTG---GAGAAGATTCCCAAGTTAAACATCCCGCCCGGCTCATTGATTAC 717
Qy 445 LeuLysHis 447
Db 718 CTCTCCTAC 726
```

RESULT 8

```
US-10-314-321A-64
; Sequence 64, Application US/10314321A
; Publication No. US20030190648A1
; GENERAL INFORMATION:
; APPLICANT: Hitachi, Ltd.
; TITLE OF INVENTION: Gene Predicting Method
```

```
FILE REFERENCE: 310101185US1
CURRENT APPLICATION NUMBER: US/10/314,321A
CURRENT FILING DATE: 2002-12-09
PRIOR APPLICATION NUMBER: JP 2002-103333
PRIOR FILING DATE: 2002-04-05
NUMBER OF SEQ ID NOS: 65
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 64
LENGTH: 1251
TYPE: DNA
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: partial sequence of AL365356
US-10-314-321A-64

Alignment Scores:
Pred. No.: 3,79e-24 Length: 1251
Score: 285.00 Matches: 85
Percent Similarity: 47.15% Conservative: 39
Best Local Similarity: 32.32% Mismatches: 96
Query Match: 11.31% Indels: 44
DB: 12 Gaps: 10

US-09-941-831A-20 (1-449) x US-10-314-321A-64 (1-1251)

QY 197 SerArgSerGlyTrpGlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHis 216
Db 65 ACCAGGTACCGTGG--ACTCATC-ACGCCCTGCAGCGCAGTCTGCGGGCCAG 120

QY 217 LeuSerCysLeuGlnValLeuAlaHisGlyAlaAspValAspSerLeuAspVallys 236
Db 121 GCGCGGTGTGTGAGTGTGTGCGGCTGCGGCGCCAGTGTGATGCTCGCAACATCGAC 180

QY 237 AlaGlnThrProLeuPheThrAlaValSerHisGlyHisLeuAspCysValargValLeu 256
Db 181 GGCAGCACCCCGCTCTGCGATCGCTCGCGCATCATCGAGTGTGTGAAGCTCTTG 240

QY 257 LeuGluAlaGlyAlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThr 276
Db 241 CTGTCTCAGCGGCCNAGTCAACCTCCCTCTGTAC---ACAGCGTCCCTCCCTGCAGCAG 297

QY 277 AlaAlaArgAspGlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAla 296
Db 298 GCCTGCATCAGCGGGAGTTCGGAATGTGTGAGGCTTCTTATTGACGTGCGG-----GCC 351

QY 297 AnValLysAlaLysLeuProValTrpAlaSerAsnIleAlaSerCys-----SerGly 314
Db 352 AATCTGGAAGCGCAC-----GATTGCCATTTTGGGACC 384

QY 315 ProLeuTyrLeuAlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHis 334
Db 385 CTCTGCGAGTGTGCTGTCGCCGGGAGCATCTGAGCTGTCTAAGTGTGCTCAATGCA 444

QY 335 GlyAlaAspProAspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgValProArgPro 354
Db 445 GGGGCCAACGTGAAT-----GCG 462

QY 355 ArgThrLeuLeuGluLeuCysLeuHisAsnCysGlu-----ProGluTyrIle 371
Db 463 GCAAAAGCTTCATGAGACTGCCCTTACCACCGCGGCCAAGGTCAAGAAATGTTGACCTCATC 522

QY 372 GlnLeuLeuLeuAspPheGlyAlaAlaIleTyrLeu-----Pro 384
Db 523 GAGATGCTTATCGAGTTTGGCGCAACATCTAGCCCGGGACACACCGCGGGAGAACGCCG 582

QY 385 SerLeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeuLeuGlnAla 404
Db 583 TCT-----GACTACAGTGGACGACGCGCTCCCGCCAGTGTCTCGAGTACTAC 633

QY 405 ArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgAlaLeuCys 424
Db 634 GAAAGACACCTCTGACTCTGTACAGCTCTGCGAGGCTGAACCTTGAGGAAGGCCACTGCG 693
```

```
QY 425 GlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMetLeuLeuSerTyr 444
Db 694 GTCCGAGGGCTG--GAGAAGATTCCCAAGTTAAACATCCCGCCCGGCTCATTTGATTAC 750

QY 445 LeuLysHis 447
Db 751 CTCTCCTAC 759

RESULT 9
US-09-919-039-133
Sequence 133, Application US/09919039
Publication No. US20030108871A1
GENERAL INFORMATION:
APPLICANT: Kaser, Matthew R.
TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
FILE REFERENCE: PA-0035 US
CURRENT APPLICATION NUMBER: US/09/919,039
CURRENT FILING DATE: 2002-09-09
PRIOR APPLICATION NUMBER: 60/222,113
PRIOR FILING DATE: 2000-07-28
NUMBER OF SEQ ID NOS: 401
SOFTWARE: PERL Program
SEQ ID NO 133
LENGTH: 3482
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030108871A1 1823159CBI
US-09-919-039-133

Alignment Scores:
Pred. No.: 2,93e-23 Length: 3482
Score: 264.00 Matches: 80
Percent Similarity: 42.00% Conservative: 46
Best Local Similarity: 26.67% Mismatches: 106
Query Match: 11.27% Indels: 68
DB: 11 Gaps: 12

US-09-941-831A-20 (1-449) x US-09-919-039-133 (1-3482)

QY 5 GlyTyrLysLeuLeuAspProIleserAspPheGlyArgMetGlyIleProAsnArgAsn 24
Db 600 GGATGCAGACAGTTTACATCCAGTGGGAAGAAATACAGGAGGCGAGGCTTGCCCAATCACCAT 659

QY 25 TrpThrIleThrAspAlaAsnArgAsnTyrGluIleCysSerThrTyrProGluIle 44
Db 660 TGGAGAATAAATCTTTTATTAAAGTGTCTATGAGCTCTGTGACACTTACCTCTCTCTTTG 719

QY 45 ValValProLysSerValThrLeuGlyThrValGlySerSerLysPheArgSerLys 64
Db 720 GTGGTTCCGTATCGTGGCTTCAGATGATGACCTCCGAGAGTTGCAACTTTTAGTCCCGA 779

QY 65 GluArgValProValLeuSerTyrLeuTyrLysGluAsnAsnAlaAlaIleCysArgCys 84
Db 780 AATCGAATTCAGTGTCTGTCATGGATTTCATCCAGAAAATAAGACGGTCTATTGTGCGCTGC 839

QY 85 SerGlnProLeuSerGlyPheTyr---ThrArgCysValAspAspGluLeuLeuGlu 103
Db 840 AGTCAGCTCTTGTGCGTATGAGTGGGAACGAAATAAAGATGATGAGAAATATCTCGAT 899

QY 104 AlaIleSerGlnThrAsnProGlySerGlnPheMetTyrValValAspThrArgProLys 123
Db 900 GTTATCAGGAGACTTAATAACAAATTTCTAAACTCACCATTATGATCAAGACCCAGC 959

QY 124 Ile-----TrpHis 126
Db 960 GTAAATGCAGTGGCCCAACAGGCAACAGGAGGAGATATAAGATGATGATCATCAT 1019

QY 127 PheLeuValLeuLeuMetArgIleValLeuGlnLeuAlaLysMetAsnLeuMet----- 144
Db 1020 AACGCGCAACTTTTCTTC-----TTAGACATTTCAATAATATTCTATTGTCGGGAA 1070
```


; CURRENT FILING DATE: 2001-06-18
 ; PRIOR APPLICATION NUMBER: 60/212,222
 ; PRIOR FILING DATE: 2000-06-16
 ; NUMBER OF SEQ ID NOS: 3
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 972
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(972)
 ; US-09-884-870-3

Alignment Scores:
 Pred. No.: 8,27e-22 Length: 990
 Score: 246.00 Matches: 82
 Percent Similarity: 45.99% Conservative: 44
 Best Local Similarity: 29.93% Mismatches: 100
 Query Match: 10.50% Indels: 48
 DB: 9 Gaps: 9

US-09-941-831A-20 (1-449) x US-09-884-870-3 (1-990)

QY 206 ThrProLeuArgLeuAlaalaSerTyGlyHisLeuSerCysLeuGlnValLeuLeuAla 225
 DB 199 TCCCCACTTCATGAAGCTCAGCTCAGGGCGCTTACTGGCCCTTAAACCTTTAATTGCA 258
 QY 226 HisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPheThrAlaVal 245
 DB 259 CAAGGTGTCAATGTGAACCTTGTGACAAATTAACCGGGTGTCTCTCTCCACGAGGCATGC 318
 QY 246 SerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerProGlyGly 265
 DB 319 CTTGGAGGTGACGTGGCTGTGCCAAGCCTTATTGGAATGTTGGTGCACACGTCATGGA 378
 QY 266 SerIleTyAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaValAlaIle 285
 DB 379 GTGACAGTTCACGGACCCACCCCTTCAATGCTTGTGTCAGCGGCGAGTGTGATGT 438
 QY 286 LeuGlnGluLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeuProValTrp 305
 DB 439 GTCAATGTGTCTGTGAGTTCGGAGCAAGGCCAGTTCGAGGTGCACCTG----- 489
 QY 306 AlaSerAsnIleAlaSerCysSerGlyProLeuTyLeuAlaAlaValTyGlyHisLeu 325
 DB 490 -----GCCCTCCCATCCATGAGCGAGTGAAGAGAGGTGCACAGA 528
 QY 326 AspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyAsn----- 341
 DB 529 GAGTGTGATGAGATCTGTGTCGAATATATGTTAATGACCATGAGGTGCTCAGCTC 588
 QY 342 -----CysThrAspGlnGlyLeuLeuAlaArgValProArgPro 354
 DB 589 GGAACCTCCCTATATGTGGCTGCACCTACCAG-----AGGTAGACTGTGTG 636
 QY 355 ArgThrLeuLeuLeuLeu-----CysLeu 362
 DB 637 AGAAACTTCTAGAAATTAGAGCCAGTGTGCACCATGCGGCGTGGTGCACACCCCACTC 696
 QY 363 His-----HisAsnCysGluProGluTyIleGlnLeuLeuIleAspPheGlyAla 379
 DB 697 CATGTGTCAGCGGCGAGTCCATGTGGAGGTCACTCACCTGTACCACTATGAGCT 756
 QY 380 AsnIleTyLeuProSerLeu-----SerLeuAspLeuThrSer---GlnAsp 394
 DB 757 AACCTGAAGCGTAGAAATGTCTAGGCAAAAGTGCCTTATCTGCGCGCTCCAAAAGC 816
 QY 395 AspLysGlyIleAlaLeuLeuGlnAlaArgAlaThrProArgSerLeuLeuSerGln 414
 DB 817 AGCGTGGAGCAGCGACTCTTGCTC-----CGTGAAGGCCCACTCTCTCTTTCCCACTC 870
 QY 415 ValArgLeuValValArgAlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGln 434

DB 871 TGCCGCTGTGTGTCCGGAAGTGTCTCGTGAGCATGTCTAT---CAAGCCATCCACAG 927
 QY 435 LeuAspIleProProMetLeuIleSerTyLeuLysHisGln 448
 DB 928 CTACATCTGCCAGAGCCCACTCGACGATCTCTCTTATACCAA 969

RESULT 12

US-09-884-870-1
 ; Sequence 1, Application US/09884870
 ; Patent No. US20020065223A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Glucksmann, Maria A.
 ; APPLICANT: Kadambi, Vivek
 ; TITLE OF INVENTION: 33358, A NOVEL HUMAN ANKYRIN FAMILY MEMBER AND USES THEREOF
 ; FILE REFERENCE: MNI-162CP
 ; CURRENT APPLICATION NUMBER: US/09/884,870
 ; CURRENT FILING DATE: 2001-06-18
 ; PRIOR APPLICATION NUMBER: 60/212,222
 ; PRIOR FILING DATE: 2000-06-16
 ; NUMBER OF SEQ ID NOS: 3
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1
 ; LENGTH: 1538
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (75)...(1046)
 ; US-09-884-870-1

Alignment Scores:
 Pred. No.: 1,75e-21 Length: 1538
 Score: 246.00 Matches: 82
 Percent Similarity: 45.99% Conservative: 44
 Best Local Similarity: 29.93% Mismatches: 100
 Query Match: 10.50% Indels: 48
 DB: 9 Gaps: 9

US-09-941-831A-20 (1-449) x US-09-884-870-1 (1-1538)

QY 206 ThrProLeuArgLeuAlaalaSerTyGlyHisLeuSerCysLeuGlnValLeuLeuAla 225
 DB 273 TCCCCACTTCATGAAGCTCAGCTCAGGGCGCTTACTGGCCCTTAAACCTTTAATTGCA 332
 QY 226 HisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPheThrAlaVal 245
 DB 333 CAAGGTGTCAATGTGAACCTTGTGACAAATTAACCGGGTGTCTCTCTCCACGAGGCATGC 392
 QY 246 SerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerProGlyGly 265
 DB 393 CTTGGAGGTGACGTGGCTGTGCCAAGCCTTATTGGAATGTTGGTGCACACGTCATGGA 452
 QY 266 SerIleTyAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaValAlaIle 285
 DB 453 GTGACAGTTCAGGAGCCACACCCCTTCAATGCTTGTGTCAGCGGCGAGTGTGATGT 512
 QY 286 LeuGlnGluLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeuProValTrp 305
 DB 513 GTCAATGTGTCTGTGAGTTCGGAGCAAGGCCAGTTCGAGGTGCACCTG----- 563
 QY 306 AlaSerAsnIleAlaSerCysSerGlyProLeuTyLeuAlaAlaValTyGlyHisLeu 325
 DB 564 -----GCCCTCCCATCCATGAGCGAGTGAAGAGAGGTGCACAGA 602
 QY 326 AspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyAsn----- 341
 DB 603 GAGTGTGATGAGATCTCTGTGCAATATATGTTAATGACCATGAGGTGCTCAGCTC 662
 QY 342 -----CysThrAspGlnGlyLeuLeuAlaArgValProArgPro 354
 DB 663 GGAACCTCCCTATATGTGGCTGCACCTACCAG-----AGGTAGACTGTGTG 710

Db 311 AGTCAGCCTCTTTCGGTATGATGGGAAACGAATAAAGATGATGAGAAATATCTCGAT 370
QY 104 AlaiSerGlnThrAsn 109
Db 371 GTTATCAGGGAGACTAAT 388

RESULT 15

US-10-025-380-28
; Sequence 28, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick Thomas S.
; APPLICANT: Carter, Darrick

; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS

; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE

; FILE REFERENCE: 210121.471C14

; CURRENT APPLICATION NUMBER: US/10/025,380

; CURRENT FILING DATE: 2001-12-19

; NUMBER OF SEQ ID NOS: 1129

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 28

; LENGTH: 401

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 212

; OTHER INFORMATION: n = A,T,C or G

US-10-025-380-28

Alignment Scores:
Pred. No.: 2,8e-22 Length: 401
Score: 244.50 Matches: 46
Percent Similarity: 62.26% Conservative: 20
Best Local Similarity: 43.40% Mismatches: 39
Query Match: 10.44% Indels: 1
DB: 13 Gaps: 1

US-09-941-831A-20 (1-449) x US-10-025-380-28 (1-401)

QY 5 GlyTyrLysLeuIleAspProIleSerAspPheGlyArgMetGlyIleProAsnArgAsn 24
Db 71 GGATGGACAGTTTACATCCAGTGGAGATACAGGAGGACGGCTTGCCCAATACCAT 130
QY 25 TrpThrIleThrAspAlaAsnArgAsnTyrGluIleCysSerThrTyrProGluIle 44
Db 131 TGGAGAATAAATTTTATTAAAGTGCTATGAGCTCTGCGACACTTACCCCTGCTCTTTG 190
QY 45 ValValProLysSerValThrLeuGlyThrValValGlySerSerLysPheArgSerLys 64
Db 191 GTGGTTCCGATATCGTCCCTCANATGATGACCTCCGGAGAGTTGCAACTTTTAGTCCCGA 250
QY 65 GluArgValProValLeuSerTyrLeuTyrLysGluAsnAsnAlaIleCysArgCys 84
Db 251 AATCGAATTCAGTGTCTGTCAGATTCATCCAGAAAATAAGACGGTCAATTGTGGCTCC 310
QY 85 SerGlnProLeuSerGlyPheTyr---ThrArgCysValAspAspGluLeuLeuGlu 103
Db 311 AGTCAGCCTCTTTCGGTATGATGGGAAACGAATAAAGATGATGAGAAATATCTCGAT 370

QY 104 AlaiSerGlnThrAsn 109
Db 371 GTTATCAGGGAGACTAAT 388

Search completed: December 1, 2003, 10:29:40
Job time : 2308 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2003, 07:34:56 ; Search time 44 Seconds
(without alignments)
981.358 Million cell updates/sec

Title: US-09-941-831A-20
Perfect score: 2343
Sequence: 1 MRESGWKLIDPISDFGRMGI.....QAINQLDIPPLMISYLKHQL 449

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	313.5	13.4	676	2 T22532	hypothetical prote
2	224.5	9.6	1848	2 S37771	ankyrin, erythrocy
3	224.5	9.6	1862	2 I49502	ankyrin - mouse
4	224.5	9.6	1881	1 SJHUK	ankyrin 1, erythro
5	221.5	9.5	1856	2 B35049	ankyrin 1, erythro
6	221.5	9.5	1880	2 A35049	ankyrin 1, erythro
7	218.5	9.3	2039	2 T15347	ankyrin-related un
8	214.5	9.2	1549	2 T13940	ankyrin - fruit fl
9	212.5	9.1	559	2 T37958	probable myotubula
10	212	9.0	1062	2 T14151	inv protein - mous
11	210.5	9.0	688	2 S57131	hypothetical prote
12	208	8.9	1062	2 T30255	inversin - mouse
13	206.5	8.8	1765	2 T42714	ankyrin 3, splice
14	206.5	8.8	1940	2 T42715	ankyrin 3, splice
15	206.5	8.8	1943	2 T42713	ankyrin 3, splice
16	206.5	8.8	1961	2 T42716	ankyrin 3, splice
17	205	8.7	276	2 T12477	hypothetical prote
18	205	8.7	4377	2 A55575	ankyrin 3, long sp
19	203.5	8.7	1435	2 T32930	hypothetical prote
20	202.5	8.6	3824	2 S37431	ankyrin 2, neurona
21	199.5	8.5	557	2 T46507	hypothetical prote
22	196.5	8.4	426	2 AE2149	hypothetical prote
23	196.5	8.4	1031	2 T43458	hypothetical prote
24	191	8.2	368	2 T18184	ankyrin repeat pro
25	187	8.0	410	1 WWVZF7	BamH1-ORF7 protein
26	181	7.7	799	2 T48451	myotubularin-like
27	181	7.7	845	2 T07052	probable potassium
28	180.5	7.7	791	2 T42691	hypothetical prote
29	177.5	7.6	426	2 E84689	probable RING zinc

30	177	7.6	887	2 T03939	potassium channel
31	174.5	7.4	888	2 D84650	probable potassium
32	171.5	7.3	2584	2 T24158	hypothetical prote
33	171.5	7.3	2606	2 T24157	hypothetical prote
34	169.5	7.2	838	2 S23606	potassium channel
35	169.5	7.2	857	2 S62694	potassium channel
36	169	7.2	262	2 T17495	ankyrin repeat pro
37	168.5	7.2	679	2 B45771	2-5A-dependent RNA
38	168	7.2	247	2 D84448	probable ankyrin I
39	167	7.1	323	2 B47169	ankyrin-like repea
40	166	7.1	1398	2 T21884	hypothetical prote
41	165	7.0	397	2 T46445	hypothetical prote
42	164.5	7.0	1066	2 S72479	Nuc-2 protein - Ne
43	164	7.0	807	2 T12177	potassium channel
44	161	6.9	180	2 A75579	ankyrin-related pr
45	160.5	6.9	1411	2 S30355	alpha-latroinsecto

ALIGNMENTS

RESULT 1

T22532
hypothetical protein F53A2.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22532
R:Lloyd, C.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19576
A:Accession: T22532
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-676 <WIL>
A:Cross-references: EMBL:Z81546; PIDN: CAB04456.1; GSPDB: GN00021; CESP: F53A2.8
A:Experimental source: clone F53A2
C:Genetics:
A:Gene: CESP:F53A2.8
A:Map position: 3
A:Introns: 10/3; 78/3; 104/1; 419/3; 449/3; 477/3; 652/3

Query Match	13.4%	Score	313.5	DB 2	Length	676			
Best Local Similarity	35.3%	Pred. No.	9.6e-18						
Matches	59	Conservative	37	Mismatches	48	Indels	23	Gaps	2
QY	5	GWKLIDPISDFGRMGIPNRNWTITDANRYEICSTY	PEIIVVPKSVTLGTGVGSSKFRSK	64					
Db	125	GWRLDWNSEMTROGITKQTSQWTESINEGYTICDTPY	PNKLPWPTAASTSVLLGSKCFRGR	184					
QY	65	ERVPLSYLYKENNAI	CRCQPLSGFYTRCVDDLELLEAISQTNPGSQFMVVDTRPKI	124					
Db	185	GRLPVLTTFHQOTEAL	CAQCPLTGFSAFCVEDEKLMELVGKANTNSDNLFLVDTRPKY	244					
QY	125	-----	-----	-----	-----	-----	-----	-----	-----
Db	245	NAMVNKVGKGFEDERNYSNMRFHFDIENIHVMRASQ	ARLLDAVTK	291					

RESULT 2

S37771
ankyrin, erythrocyte - mouse
C:Species: Mus musculus (house mouse)
C>Date: 27-May-1994 #sequence_revision 03-Aug-1995 #text_change 13-Aug-1999
C:Accession: S37771
J.Birkmeier, C.S.; White, R.A.; Peters, L.L.; Hall, E.J.; Lux, S.E.; Barker, J.E.
J. Biol. Chem. 268, 9533-9540, 1993
A:Title: Complex patterns of sequence variation and multiple 5' and 3' ends are found am
A:Reference number: S37771; MUID: 93252825; PMID: 8486643
A:Accession: S37771
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1848 <BIR>
A:Cross-references: EMBL:X69063; NID: g311816; PIDN: CAA48801.1; PID: g311817

C:Superfamily: ankyrin; ankyrin repeat homology
 C:Keywords: alternative splicing
 F:48-80/Domain: ankyrin repeat homology <AN01>
 F:81-113/Domain: ankyrin repeat homology <AN02>
 F:114-146/Domain: ankyrin repeat homology <AN03>
 F:147-175/Domain: ankyrin repeat homology <AN04>
 F:176-208/Domain: ankyrin repeat homology <AN05>
 F:209-241/Domain: ankyrin repeat homology <AN06>
 F:242-274/Domain: ankyrin repeat homology <AN07>
 F:275-307/Domain: ankyrin repeat homology <AN08>
 F:308-340/Domain: ankyrin repeat homology <AN09>
 F:341-373/Domain: ankyrin repeat homology <AN10>
 F:374-406/Domain: ankyrin repeat homology <AN11>
 F:407-439/Domain: ankyrin repeat homology <AN12>
 F:440-472/Domain: ankyrin repeat homology <AN13>
 F:473-505/Domain: ankyrin repeat homology <AN14>
 F:506-538/Domain: ankyrin repeat homology <AN15>
 F:539-571/Domain: ankyrin repeat homology <AN16>
 F:572-604/Domain: ankyrin repeat homology <AN17>
 F:605-637/Domain: ankyrin repeat homology <AN18>
 F:638-670/Domain: ankyrin repeat homology <AN19>
 F:671-703/Domain: ankyrin repeat homology <AN20>
 F:704-736/Domain: ankyrin repeat homology <AN21>
 F:737-769/Domain: ankyrin repeat homology <AN22>
 F:770-802/Domain: ankyrin repeat homology <AN23>

Query Match 9.6%; Score 224.5; DB 2; Length 1848;
 Best Local Similarity 32.2%; Pred. No. 8.6e-10;
 Matches 76; Conservative 37; Mismatches 86; Indels 37; Gaps 7;
 QY 169 ALNQAAYVNDSTLDQLRQRYKRFINSRSGWVGTPPLRLAASVGHLSCLQVLLAHGA 228
 DB 180 ALHIAARNDDTTRTAAVLLQNDPNPDLV-SKTGF-----TPLHIAAHYENLVNAQLLNRCGA 234
 QY 229 DVDSLDVKAQTPLFTAVSHGHLDVRYLLEAGASPGGSIYNNCSPLVLTAAARDGAVAILQE 288
 DB 235 SVNFTPQNGITPLHIAASRGVIMVRLLLDRGAQIETRTKDELTPHCAARNGHVRISEI 294
 QY 289 LLDHGAEANVKAFLPWASNIASCSGPLYLAAYVGHLDVCFRLLLLHGADPDYNCNTD---- 344
 DB 295 LLDHGAPIQAKTK-----NGLSPIHMAAQGDHLDVRLLLQYNAEIDITLDHLTP 345
 QY 345 -----QGLLARV-----PRPTLLEIC-LHHNCEPEYI---QLLIDFGANI 381
 DB 346 LHVAACHGHRHVRVAKVLLDKGAKPNSRALNGFTPLHIAACKNHRVWELLKTKGASI 401

RESULT 3
 I49502
 ankyrin - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
 C:Accession: I49502
 R:White, R.A.; Birkenmeier, C.S.; Peters, L.L.; Barker, J.E.; Lux, S.E.
 Mamm. Genome 3, 281-285, 1992
 A>Title: Murine erythrocyte ankyrin cDNA: Highly conserved regions of the regulatory domain
 A:Reference number: I49502; MUID:192345717; PMID:1386265
 A:Status: preliminary
 A:Accession: I49502
 A:Molecule type: mRNA
 A:Residues: 1-1862 <RES>
 A:Cross-references: GB:M84756; NID:gl91939; PIDN:AAA37236.1; PID:gl91940
 C:Genetics:
 A:Gene: Ank-1
 C:Superfamily: ankyrin; ankyrin repeat homology
 C:Keywords: alternative splicing
 F:40-72/Domain: ankyrin repeat homology <AN01>
 F:73-105/Domain: ankyrin repeat homology <AN02>
 F:106-138/Domain: ankyrin repeat homology <AN03>
 F:139-167/Domain: ankyrin repeat homology <AN04>
 F:168-200/Domain: ankyrin repeat homology <AN05>
 F:201-233/Domain: ankyrin repeat homology <AN06>
 F:234-266/Domain: ankyrin repeat homology <AN07>

F:267-299/Domain: ankyrin repeat homology <AN08>
 F:300-332/Domain: ankyrin repeat homology <AN09>
 F:333-365/Domain: ankyrin repeat homology <AN10>
 F:366-398/Domain: ankyrin repeat homology <AN11>
 F:399-431/Domain: ankyrin repeat homology <AN12>
 F:432-464/Domain: ankyrin repeat homology <AN13>
 F:465-497/Domain: ankyrin repeat homology <AN14>
 F:498-530/Domain: ankyrin repeat homology <AN15>
 F:531-563/Domain: ankyrin repeat homology <AN16>
 F:564-596/Domain: ankyrin repeat homology <AN17>
 F:597-629/Domain: ankyrin repeat homology <AN18>
 F:630-662/Domain: ankyrin repeat homology <AN19>
 F:663-695/Domain: ankyrin repeat homology <AN20>
 F:696-728/Domain: ankyrin repeat homology <AN21>
 F:729-761/Domain: ankyrin repeat homology <AN22>
 F:762-794/Domain: ankyrin repeat homology <AN23>

Query Match 9.6%; Score 224.5; DB 2; Length 1862;
 Best Local Similarity 32.2%; Pred. No. 8.7e-10;
 Matches 76; Conservative 37; Mismatches 86; Indels 37; Gaps 7;
 QY 169 ALNQAAYVNDSTLDQLRQRYKRFINSRSGWVGTPPLRLAASVGHLSCLQVLLAHGA 228
 DB 172 ALHIAARNDDTTRTAAVLLQNDPNPDLV-SKTGF-----TPLHIAAHYENLVNAQLLNRCGA 226
 QY 229 DVDSLDVKAQTPLFTAVSHGHLDVRYLLEAGASPGGSIYNNCSPLVLTAAARDGAVAILQE 288
 DB 227 SVNFTPQNGITPLHIAASRGVIMVRLLLDRGAQIETRTKDELTPHCAARNGHVRISEI 286
 QY 289 LLDHGAEANVKAFLPWASNIASCSGPLYLAAYVGHLDVCFRLLLLHGADPDYNCNTD---- 344
 DB 287 LLDHGAPIQAKTK-----NGLSPIHMAAQGDHLDVRLLLQYNAEIDITLDHLTP 337
 QY 345 -----QGLLARV-----PRPTLLEIC-LHHNCEPEYI---QLLIDFGANI 381
 DB 338 LHVAACHGHRHVRVAKVLLDKGAKPNSRALNGFTPLHIAACKNHRVWELLKTKGASI 393

RESULT 4
 SJHUK
 ankyrin 1, erythrocyte splice form 1 - human
 N:Alternate names: ankyrin 2.1, erythrocyte; ankyrin-R
 N:Contains: ankyrin 2.2
 C:Species: Homo sapiens (man)
 C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 22-Jun-1999
 C:Accession: S08275; A33219; PC2220; A35443
 R:Lux, S.E.; John, K.M.; Bennett, V.
 Nature 344, 36-42, 1990
 A>Title: Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure w
 A:Reference number: S08275; MUID:90150830; PMID:2137557
 A:Accession: S08275
 A:Molecule type: mRNA
 A:Residues: 1-1881 <LU1>
 A:Cross-references: EMBL:X16609; NID:g28701; PIDN:CAA34610.1; PID:g28702
 A:Accession: A33219
 A:Molecule type: protein
 A:Residues: 2-7, X, 9-17, X, 19-20, T, 22-30, 733-749, A, 751-753, 828-833, X, 835-855, X,
 X, 1367, 1383-1427, 1601-1630, 1686-1698, D, 1700, 1763-1772 <LUX>
 A>Note: 845-Arg and 1392-Thr were also found
 R:Hermann, J.; Barel, M.; Frade, R.
 Biochem. Biophys. Res. Commun. 204, 453-460, 1994
 A>Title: Human erythrocyte ankyrin, a cytoskeleton component, generates the p57 membrane
 A:Reference number: PC2220; MUID:95071348; PMID:7526850
 A:Accession: PC2220
 A:Molecule type: protein
 A:Residues: 910-929 <HER>
 R:David, L.H.; Bennett, V.
 J. Biol. Chem. 265, 10589-10596, 1990
 A>Title: Mapping the binding sites of human erythrocyte ankyrin for the anion exchanger
 A:Reference number: A35443; MUID:90285190; PMID:2141335
 A:Accession: A35443
 A:Molecule type: protein
 A:Residues: X, 5, X, 7-12, 403-417, X, 419-422, H, 424, LQ, 797-800, L, 802-814, 862-863, X,

C:Genetics: A:Gene: GDB:ANK1; ANK
 A:Accession: B35049
 A:Cross-references: GDB:118737; OMIM:182900
 A:Map position: 8p11.2-8p11.2
 A:Superfamily: ankryrin; ankryrin repeat homology
 C:Keywords: alternative splicing; phosphoprotein
 F:2-1881/Product: ankryrin 1, erythrocyte form 1 #status predicted <MAT1>
 F:2-1512.1675-1881/Product: ankryrin 2.2, erythrocyte #status predicted <MAT2>
 F:2-827/Domain: 89K #status predicted <DOM1>
 F:2-827/Region: anion exchange protein binding
 F:44-76/Domain: ankryrin repeat homology <AN01>
 F:77-109/Domain: ankryrin repeat homology <AN02>
 F:110-142/Domain: ankryrin repeat homology <AN03>
 F:143-171/Domain: ankryrin repeat homology <AN04>
 F:172-204/Domain: ankryrin repeat homology <AN05>
 F:205-237/Domain: ankryrin repeat homology <AN06>
 F:238-270/Domain: ankryrin repeat homology <AN07>
 F:271-303/Domain: ankryrin repeat homology <AN08>
 F:304-336/Domain: ankryrin repeat homology <AN09>
 F:337-369/Domain: ankryrin repeat homology <AN10>
 F:370-402/Domain: ankryrin repeat homology <AN11>
 F:403-435/Domain: ankryrin repeat homology <AN12>
 F:436-468/Domain: ankryrin repeat homology <AN13>
 F:469-501/Domain: ankryrin repeat homology <AN14>
 F:502-534/Domain: ankryrin repeat homology <AN15>
 F:535-567/Domain: ankryrin repeat homology <AN16>
 F:568-600/Domain: ankryrin repeat homology <AN17>
 F:601-633/Domain: ankryrin repeat homology <AN18>
 F:634-666/Domain: ankryrin repeat homology <AN19>
 F:667-699/Domain: ankryrin repeat homology <AN20>
 F:700-732/Domain: ankryrin repeat homology <AN21>
 F:733-765/Domain: ankryrin repeat homology <AN22>
 F:766-798/Domain: ankryrin repeat homology <AN23>
 F:828-1382/Region: spectrin binding
 F:1383-1881/Domain: 55K #status predicted <DOM3>

Query Match 9.6%; Score 224.5; DB 1; Length 1881;
 Best Local Similarity 28.0%; Pred. No. 8.8e-10;
 Matches 80; Conservative 50; Mismatches 115; Indels 41; Gaps 7;
 QY 169 ALNQAAYDNDSTYLLQDLRQERYKRFINSRGWVGPTPLRLAASYGHLSCLQVLLAHGA 228
 DB 176 ALHIAARNDDTRTAAVLLQNDPNPDLV-SKTGF-----TFLHTAAHYENLNVAQLLNRCGA 230
 QY 229 DVDSLDRVQAOTPLFTAVSHGHLDGCVRVLEAGASPGSGSYNNCSPLVLTARPGAVAILQE 288
 DB 231 SYNFTPQNGITPLHTASRRGNVIMVRLLDRCQAQIETKTDELTLPLHCAARNGHVRISIEI 290
 QY 289 LLDHGAEANVKAKLPWASNIASCSGPLYLAAYVGHLDGCFRLLLLHGGADPDYNCDDQGLL 348
 DB 291 LLDHGAPTOAKTK-----NGLSPIHMAAQGHLDGCVRLLLQYDAEIDITLDH---- 338
 QY 349 ARVPRPTLLEICLHNCPEPEYIQLLIDFGANIYLPSSL-----LDLTS 392
 DB 339 ---LTPHVAACHGHH---RVAKVLLDGAKPNSRALNGFTPLHTACKNHVRVWELL 391
 QY 393 QDDKGIALLOARATPRSLLSQV-RLVVRRAICQAGQQAQINQLDI 437
 DB 392 KTGSIDAVTESGLTPLHVASFPMGHLPIVKNLLQRGASPNVSNVKV 437

RESULT 5
 B35049
 N:Alternate names: ankryrin 1, erythrocyte splice form 3 - human
 N:Alternate names: ankryrin 2.1, erythrocyte; ankryrin-R
 N:Contains: ankryrin 2.2, erythrocyte
 C:Species: Homo sapiens (man)
 C:Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 10-Jul-1998
 C:Accession: B35049
 R:Lambert, S.; Yu, H.; Prchal, J.T.; Lawler, J.; Ruff, P.; Speicher, D.; Cheung, M.C.; K
 Proc. Natl. Acad. Sci. U.S.A. 87, 1730-1734, 1990
 A:Title: cDNA sequence for human erythrocyte ankryrin.

A:Reference number: A35049; MUID:90175370; PMID:1689849
 A:Accession: B35049
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1856 <LAM>
 C:Genetics: A:Gene: GDB:ANK1; ANK
 A:Cross-references: GDB:118737; OMIM:182900
 A:Map position: 8p11.2-8p11.2
 A:Superfamily: ankryrin; ankryrin repeat homology
 C:Keywords: alternative splicing
 F:2-1856/Product: ankryrin 1, erythrocyte form 3 #status predicted <MAT>
 F:2-1513.1676-1856/Product: ankryrin 2.2, erythrocyte #status predicted <MA2>
 F:44-76/Domain: ankryrin repeat homology <AN01>
 F:77-109/Domain: ankryrin repeat homology <AN02>
 F:110-142/Domain: ankryrin repeat homology <AN03>
 F:143-171/Domain: ankryrin repeat homology <AN04>
 F:172-204/Domain: ankryrin repeat homology <AN05>
 F:205-237/Domain: ankryrin repeat homology <AN06>
 F:238-270/Domain: ankryrin repeat homology <AN07>
 F:271-303/Domain: ankryrin repeat homology <AN08>
 F:304-336/Domain: ankryrin repeat homology <AN09>
 F:337-369/Domain: ankryrin repeat homology <AN10>
 F:370-402/Domain: ankryrin repeat homology <AN11>
 F:403-435/Domain: ankryrin repeat homology <AN12>
 F:436-468/Domain: ankryrin repeat homology <AN13>
 F:469-501/Domain: ankryrin repeat homology <AN14>
 F:502-534/Domain: ankryrin repeat homology <AN15>
 F:535-567/Domain: ankryrin repeat homology <AN16>
 F:568-600/Domain: ankryrin repeat homology <AN17>
 F:601-633/Domain: ankryrin repeat homology <AN18>
 F:634-666/Domain: ankryrin repeat homology <AN19>
 F:667-699/Domain: ankryrin repeat homology <AN20>
 F:700-732/Domain: ankryrin repeat homology <AN21>
 F:733-765/Domain: ankryrin repeat homology <AN22>
 F:766-798/Domain: ankryrin repeat homology <AN23>

Query Match 9.5%; Score 221.5; DB 2; Length 1856;
 Best Local Similarity 27.6%; Pred. No. 1.5e-09;
 Matches 79; Conservative 51; Mismatches 115; Indels 41; Gaps 7;
 QY 169 ALNQAAYDNDSTYLLQDLRQERYKRFINSRGWVGPTPLRLAASYGHLSCLQVLLAHGA 228
 DB 176 ALHIAARNDDTRTAAVLLQNDPNPDLV-SKTGF-----TFLHTAAHYENLNVAQLLNRCGS 230
 QY 229 DVDSLDRVQAOTPLFTAVSHGHLDGCVRVLEAGASPGSGSYNNCSPLVLTARPGAVAILQE 288
 DB 231 SYNFTPQNGITPLHTASRRGNVIMVRLLDRCQAQIETKTDELTLPLHCAARNGHVRISIEI 290
 QY 289 LLDHGAEANVKAKLPWASNIASCSGPLYLAAYVGHLDGCFRLLLLHGGADPDYNCDDQGLL 348
 DB 291 LLDHGAPTOAKTK-----NGLSPIHMAAQGHLDGCVRLLLQYDAEIDITLDH---- 338
 QY 349 ARVPRPTLLEICLHNCPEPEYIQLLIDFGANIYLPSSL-----LDLTS 392
 DB 339 ---LTPHVAACHGHH---RVAKVLLDGAKPNSRALNGFTPLHTACKNHVRVWELL 391
 QY 393 QDDKGIALLOARATPRSLLSQV-RLVVRRAICQAGQQAQINQLDI 437
 DB 392 KTGSIDAVTESGLTPLHVASFPMGHLPIVKNLLQRGASPNVSNVKV 437

RESULT 6
 A35049
 N:Alternate names: ankryrin 1, erythrocyte splice form 2 - human
 N:Alternate names: ankryrin 2.1, erythrocyte; ankryrin-R
 N:Contains: ankryrin 2.2, erythrocyte
 C:Species: Homo sapiens (man)
 C:Date: 27-Jul-1990 #sequence_revision 01-Oct-1992 #text_change 04-Sep-1998
 C:Accession: A35049
 R:Lambert, S.; Yu, H.; Prchal, J.T.; Lawler, J.; Ruff, P.; Speicher, D.; Cheung, M.C.; K
 Proc. Natl. Acad. Sci. U.S.A. 87, 1730-1734, 1990
 A:Title: cDNA sequence for human erythrocyte ankryrin.

A:Reference number: A35049; MUID:90175370; PMID:1689849

A:Accession: A35049

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1880 <LAW>

A:Cross-references: GB:M2880

C:Genetics:

A:Gene: GDB:ANK1; ANK

A:Cross-references: GDB:118737; OMIM:182900

A:Map position: 8p11.2-8p11.2

C:Superfamily: ankyrin; ankyrin repeat homology

C:Keywords: alternative splicing; cytoskeleton

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

F:2-1513/Domain: ankyrin repeat homology <AN01>

RESULT 8
T13940
ankyrim - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C:Accession: T13940
R:Dubreuil, R.R.; Yu, J.
Proc. Natl. Acad. Sci. U.S.A. 91, 10285-10289, 1994
A:Title: Ankyrim and beta-spectrin accumulate independently of alpha-spectrin in Drosophila
A:Reference number: Z17820; PMID:95024098; PMID:7937942
A:Accession: T13940
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1549 <DUB>
A:Cross-references: EMBL:L35601; NID:G557083; PID:G557084; PIDN:AAC37208.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0011747

Query Match 9.2%; Score 214.5; DB 2; Length 1549;
Best Local Similarity 25.2%; Pred. No. 4.5e-09;
Matches 84; Conservative 53; Mismatches 121; Indels 75; Gaps 10;

QY 131 IMRIVQLAKMN-----LMDITKIFSLQDPKEEDTDTEEQALNQAVY 175
DB 480 IIRILRSKAVDAIVREGOTPLHVASRLGNIINIIMLLQHGAEINQAQNDKYSALHIAAK 539
QY 176 DNDSTYLDQLRQERYKRFINSRSGVPGTPLRLAASVGHLSCLQVILAHGADVDSLDV 235
DB 540 EQENIVQVLENGAENNAV-TKKGF-----TPLHLACKYQKQVQVQLLQNGASIDFQK 594
QY 236 KAQTPLFTAVSHGHLCVVRVLEAGASPGSIYNNCSPLVLTAAARDGAVAILQELLDHGAE 295
DB 595 NDVTPPLHVATHYNNPSIVLELLKNGSPNLCARNGQCAITHACKNYLAIAMQLLQHGAD 654
QY 296 ANVKAKPLVWASNIASCSCG-----PLYLAAYVGHLDCCFRLLE 332
DB 655 VNIISKSGFPLHLAAGGVNDVQVLLLEVYGVISAANKGLTFLHVAQEGHVLVQIILL 714
QY 333 LHGADPDYNTDQGLLARVPRPTLLEICLHNCEPEYIQLL-----IDFGANI-Y 382
DB 715 EHGAN-----ISERTNGYVTPHMAHYG-----HLDLVKFFIENDADIEMSSNIGY 761
QY 383 LPSLSLDLTSQDDKG-----IALQLQARATPRSL 411
DB 762 TP-----LHQAQOQHIMIINLLRHKAMPNAL 789

RESULT 9
T37958
probable myotubularin-like protein-tyrosine phosphatase - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T37958
R:Skelton, J.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A:Reference number: Z21756
A:Accession: T37958
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-559 <SKE>
A:Cross-references: EMBL:Z98974; PIDN:CAB11639.1; GSPDB:GN00066; SPDB:SPAC19A8.03
A:Experimental source: strain 972h-; cosmid c19A8
C:Genetics:
A:Gene: SPDB:SPAC19A8.03
A:Map position: 1
A:Introns: 8/3; 200/3

Query Match 9.1%; Score 212.5; DB 2; Length 559;
Best Local Similarity 39.2%; Pred. No. 1.7e-09;
Matches 51; Conservative 20; Mismatches 46; Indels 13; Gaps 5;

QY 4 SGWKLIDPISDFGRMGIPNR-----NWTIDANRNYEICSTYPTPEIVPKSVTLGTV 55

DB 119 SSWKSFLLNEVRRMGVGDSTQADGAGGNWRTIKINENYSECHYPQALAVPASISDSVI 178
QY 56 VGSSKFRSKERVPVLSYLKKNNAICRCSOPLSGF-YTRCVDDVDELLLEAISQTN--PGS 112
DB 179 YGCKYRSKNRFPPTUTYLHK-NSFSITRASQVGLVGRNSAQDEKLVEAIFATSIIPK 237
QY 113 QPMYVVDTRP 122
DB 238 ENL-IVDARP 246

RESULT 10
T14151
Inv protein - mouse
C:Species: Mus musculus (house mouse)
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14151
R:Yokoyama, T.; Mochizuki, T.
submitted to the EMBL Data Library, November 1997
A:Reference number: Z17888
A:Accession: T14151
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1062 <YOK>
A:Cross-references: EMBL:AF034860; NID:G3560522; PID:G3560523; PIDN:AAC34976.1
C:Genetics:
A:Gene: inv
A:Map position: 4

Query Match 9.0%; Score 212; DB 2; Length 1062;
Best Local Similarity 30.1%; Pred. No. 4.4e-09;
Matches 74; Conservative 40; Mismatches 98; Indels 34; Gaps 9;

QY 139 AKNNLMDITKIFSLQDPKEEDTDTEEQALNQAVYNDSTYLDQLRQERYKRFINSR 198
DB 297 AQSNFAETKVF--LQHPVSKDSDLEGRTPMWAAGKGNDDVLRMTL-SLKSDIDINMS 353
QY 199 SCWGVPGTPLRLAASVGHLSCLQVILAHGADVDSLDVKAQTPLFTAVSHGHLDCCFRLLE 258
DB 354 DKYG--GTALHAAALSGHVSTVKLLDNDQAQVADTDVWKHTPLFRACENGHRRDVIOTLIK 411
QY 259 AGASPGGSYNNCSPLVLTAAARDGAVAILQELLDHGAEANVKAKPLVWASNIASCSCG--PL 316
DB 412 GEARVDLVDDQGHSLHWAALGNADVQCILLENKINPNVQ-----DYAGRTPL 460
QY 317 YLAAYVGHLDCCFRLLELHGHADPDYNTDQGLLARVPRPTLLEICLHNCEPEY----IQL 373
DB 461 QCAAYGGYINCMAVLMMENADP--NIQDK-----EGRT-----ALHWSCNNGYLDAILK 507

QY 374 LIDFGA 379
DB 508 LLDFAA 513

RESULT 11
S57131
hypothetical protein YUR110w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein J2007
C:Species: Saccharomyces cerevisiae
C>Date: 03-Sep-1995 #sequence_revision 16-Feb-1996 #text_change 19-Apr-2002
C:Accession: S57131; S57133; B33478
R:Ramezani Rad, M.; Kirchrath, L.; Hollenberg, C.P.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S57111
A:Accession: S57131
A:Molecule type: DNA
A:Residues: 1-688 <RAM>
A:Cross-references: EMBL:Z49610; NID:G1015823; PIDN:CAA9640.1; PID:G1015824; MIPS:YUR111
R:Rose, M.; Koetter, P.; Entian, K.D.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S57132
A:Accession: S57133

A:Molecule type: DNA
A:Residues: 1-688 <ROS>
A:Cross-references: EMBL:Z49610; NID:g1015823; PIDN:CAA9640.1; PID:g1015824; MIPS:YJR11
Mol. Cell. Biol. 9, 4882-4888, 1989
A:Title: Arginine restriction induced by delta-N-(phosphonacetyl)-L-ornithine signals in
A:Reference number: A33478; MUID:90097903; PMID:2689969
A:Accession: B33478
A:Molecule type: DNA
A:Residues: 1-84, 'H', 86-88 <KIN>
A:Cross-references: GB:M30466; NID:g171311; PIDN:AAA66903.1; PID:g808832
A:Genetics:
A:Cross-references: SGD:S0003871
A:Map position: 10R

Query Match 9.0%; Score 210.5; DB 2; Length 688;
Best Local Similarity 21.5%; Pred. No. 3.3e-09;
Matches 84; Conservative 56; Mismatches 99; Indels 151; Gaps 15;

QY 6 WKLDIPISDFRGMIPNRN---WTITDANRNYEICSTYPPETVVPKSVTLGTWVGSSKF 61
DB 156 WDIYDPIKEPRRQGLSKDCTCPWLSTVNEHYEFCPTYPEKLFVPRSTSDILLKHASKF 215
QY 62 RSKERPVLVLYKENNAICRQSPISGFTY-RCVDDLELL-----EASOTNPG 111
DB 216 RSQRIPLVLYTHHKAATDCNLRSSQPLPGLINQRSIQDEKLVWESFNSFCNKDIRTK-- 273
QY 112 SQPMYVVDTRPKIWHFVLIMRIVLQAKNLMIDTKIFSLQLQDKEEDTDTFEKQALN 171
DB 274 ---HVIDAKPR-----TNALAQWALG 292
QY 172 QAYVNDSTYLDQLRQRYKRFINSRGVPGTPLRLAASGYHLSCLQLVLAHGAADV 231
DB 293 GGTENMDNY-----NFFLADNMGVD-----K 314
QY 232 SLDVKAQTPLFTAVSHGHLDVVRVLEAGSPGSIYNNCSPLVLTAAAGCAVAILQELLD 291
DB 315 SLKLPVTRFLGIDNIH-----IVSNTAAVMT-----EVIC 346
QY 292 HGAEANV-----KAKLPVMA---SNIASCSGPLYLAAYVGHLDVCRFLLLLHGAADPDY 340
DB 347 QGGDLNLPLEQNLIROKFSNWKULNTLILKSVMLKSIIFNHSN-----VLVHCSD-GW 401
QY 341 NCTDQGLLARVPRPTLLICLHNCPEY 370
DB 402 DRTSQVV-----SLLEICU-----DPFY 419

RESULT 12
T30255
inversin - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C:Accession: T30255
R:Morgan, D.; Turnpenny, L.; Goodship, J.; Dai, W.; Majumder, K.; Matthews, L.; Gardner,
Nature Genet. 20, 149-156, 1998
A:Title: Inversin, a novel gene in the vertebrate left-right axis pathway, is partially
A:Reference number: 220793; MUID:98442650; PMID:9771707
A:Accession: T30255
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1062 <MOR>
A:Cross-references: EMBL:AJ010902; NID:g3757671; PIDN:CAA09388.1; PID:g3757672
A:Experimental source: strain FBV/N
C:Genetics:
A:Gene: invs
A:Map position: 4

Query Match 8.9%; Score 208; DB 2; Length 1062;
Best Local Similarity 29.7%; Pred. No. 9.4e-09;
Matches 73; Conservative 41; Mismatches 98; Indels 34; Gaps 9;

QY 139 AKNLMIDITKIFSLQPDKEEDTDTTEKQALNAQVYNDSTYLDQLLRQRYKRFINSR 198

DB 297 AOSNFAETVKVF--LQHPSVKDDSDQEGRTSFMAAGKGNDDVLRITML-SLKSDIDINMS 353
QY 199 SGWGVPGTPLRLAASGYHLSCLQLVLAHGAADVSLDKAQTPLFTAVSHGHLDVVRVLE 258
DB 354 DRYG--GTALHAALSGHVSSTVKLLDNDQAQVADTDVMDKHTPLFRACEMGHRDVIOTLIK 411
QY 259 AGASPGSIIYNNCSPLVLTAAAGCAVAILQELLDHGAENAVKAKLPVWASNIASCSG--PL 316
DB 412 GGARVLDVDDGSHLSLHWAALGNAADVCCILLENKINPNVQ-----DYAGRTPI 460
QY 317 YLAAYVGHLDVCRFLLLLHGAADPDYNTDQGLLARVPRPTLLICLHNCPEY----IQL 373
DB 461 QSAAYGGNINCMVLMENNADP--NIQDK-----EGRT-----ALHWSNNGYLDAILK 507
QY 374 LIDFGA 379
DB 508 LIDFAA 513

RESULT 13
T42714
ankyrin 3, splice form 2 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
C:Accession: T42714
R:Peterson, L.L.; John, K.M.; Lu, F.M.; Bicher, E.M.; Higgins, A.; Vialamas, M.; Turtzo, L.
J. Cell Biol. 130, 313-330, 1995
A:Title: Ank3 (epithelial ankyrin), a widely distributed new member of the ankyrin gene f
the repeat domain.
A:Reference number: Z22237; MUID:95340633; PMID:7615634
A:Accession: T42714
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1765 <PTE>
A:Cross-references: EMBL:I40632; NID:g710548; PID:g710551; PIDN:AAB01605.1
A:Experimental source: strain C57BL/6J; kidney
C:Genetics:
A:Gene: Ank3
A:Map position: 10
A:Introns: 1587/1
C:Superfamily: ankyrin; ankyrin repeat homology
C:Keywords: alternative splicing

Query Match 8.8%; Score 206.5; DB 2; Length 1765;
Best Local Similarity 26.5%; Pred. No. 2.5e-08;
Matches 75; Conservative 38; Mismatches 93; Indels 77; Gaps 9;

QY 206 TPLRLAASGYHLSCLQLVLAHGAADVSLDKVKA----- 237
DB 418 TPIHVAAPFMGHVIVISQLMHGASPTNTNVRGETALHMAARSGQAEVVRVLYVDGQAVEA 477
QY 238 -----QTPFTAVSHGHLDVVRVLEAGSPGSIYNNCSPLVLTAAAGCAVAILQELLDH 292
DB 478 KAKDDQTPPLHISARLGKADIVQQLQOGASPNAAATTSVTPPLHAAAREGHEDVAAFLDH 537
QY 293 GAEANVKAAPVWASNIASCSG--PLYLAAYVGHLDVCRFLLLLHGAADPDYNTDQGLLAR 350
DB 538 GASLSITTK-----KGFTPLHVAAYGKLEVASLLLOKSGASPD--AAGKSGL---- 582
QY 351 VPRPTLLICLHNCPEYIQLLDIFGA-----NIYLPSSLDLTSQDDKGTALL-- 401
DB 583 -----TPLHVAHYD-NQKVALLLLDQGNASPHAAAKNGVTPPLHIAKKNQMDIATSLLEY 636
QY 402 -LQAPATPRSLLSQVRLVVRALCOAGQFQAINQLDIPMLIS 443
DB 637 GADANAVTRQGIASVHLAAQE-----GHVDMVSLILS 668

RESULT 14
T42715
ankyrin 3, splice form 3 - mouse
C:Species: Mus musculus (house mouse)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 1, 2003, 07:44:17 ; Search time 2290 Seconds
(without alignments)
4765.375 Million cell updates/sec

Title: US-09-941-831A-20
Perfect score: 2343
Sequence: 1 MRESGWKLDIPDFGRMG.....QAINQLDIPPLMSLYLKHL 449

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlpl
-MODEL=frame_p2n.model -Q=US09941831/runat_01122003_073527_7841/app_query.fasta_1.647
-DB=EST -QFMT=fastap -SUFFIX=rest -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09941831 @CGN 1.1 2874 @runat_01122003_073527_7841 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_nam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1459	62.3	1262	11	AK009644	AK009644 Mus muscu
2	1160	49.5	719	12	BM190670	BM190670 POSM01000
3	1113	47.5	644	13	EX107270	EX107270 BX107270
4	890.5	38.0	861	13	BU254916	BU254916 603747645
5	888	37.9	908	13	BU220575	BU220575 603106716
6	843.5	36.3	678	29	AG127205	AG127205 Pan trogl
7	838	35.8	849	13	BU228932	BU228932 603798270
8	805.5	34.4	713	14	CA364256	CA364256 639132 NC
9	805.5	34.4	758	13	EX073379	EX073379 BX073379
10	776	33.1	446	14	W76256	W76256 zd66f07.r1
11	756.5	32.3	755	14	CA368575	CA368575 644799 NC
12	752	32.1	518	28	AZ725513	AZ725513 RPCI-24-1
13	739.5	31.6	680	14	CA358977	CA358977 631831 NC
14	708.5	30.2	932	14	BY709303	BY709303 BY709303
15	683	29.2	643	10	BB661935	BB661935 BB661935
16	675	28.8	627	10	BB662275	BB662275 BB662275
17	673	28.7	647	14	CB516250	CB516250 ssalrpb50
18	649.5	27.7	990	14	CB207207	CB207207 AGENCOURT
19	647	27.6	1060	29	CC217390	CC217390 CH261-79C
20	640.5	27.3	878	14	CB198136	CB198136 AGENCOURT
21	615.5	26.3	690	13	BU382590	BU382590 603858336
22	608.5	26.0	776	13	BU385175	BU385175 603583214
23	601	25.7	538	12	BI290522	BI290522 UI-R-DK0-
24	585	25.0	482	9	AA633985	AA633985 ac33f04.s
25	577	24.6	1201	9	AL529411	AL529411 AL529411
26	574	24.5	372	28	AZ725227	AZ725227 RPCI-24-1
27	573.5	24.5	675	14	CA375712	CA375712 653860 NC
28	567.5	24.2	524	14	W58534	W58534 zd20c05.r1
29	560.5	23.9	583	13	BU438451	BU438451 604147108
30	544	23.2	655	9	AL776681	AL776681 AL776681
31	534	22.8	486	10	BF889720	BF889720 289285 MA
32	531	22.7	708	14	CB591162	CB591162 AGENCOURT
33	517	22.1	543	14	W72063	W72063 zd66f07.r1
34	516	22.0	567	12	BJ493353	BJ493353 BJ493353
35	480.5	20.5	1301	11	AK006515	AK006515 Mus muscu
36	480	20.5	723	14	CB593182	CB593182 AGENCOURT
37	478	20.4	631	13	BQ395305	BQ395305 NISC pg14
38	472	20.1	625	10	AW957331	AW957331 EST369521
39	461	19.7	475	9	AI546016	AI546016 fb77b07.y
40	451	19.2	859	9	AJ454588	AJ454588 AJ454588
41	450	19.2	990	12	BM545864	BM545864 AGENCOURT
42	449	19.2	611	12	BM192381	BM192381 dai22g11.
43	446	19.0	464	14	W58499	W58499 zd20c05.r1
44	446	19.0	561	12	BJ079290	BJ079290 BJ079290
45	443.5	18.9	464	10	BF555515	BF555515 UI-R-AI-d

ALIGNMENTS

RESULT 1
AK009644
LOCUS
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310036C05 product:ankyrin repeat domain-containing SOCS box protein 12, full insert sequence.
ACCESSION AK009644
VERSION AK009644.1 GI:12844562
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.


```
Db 324 GACCACCTTCTACACGAGAACGTTTATAACGGTTTCATCAACGAGGAGTGGCTGGGT 383
Qy 203 ValProGlyThrProLeuArgLeuAlaSerTyrglyHisLeuSerCysLeuGlnVal 222
Db 384 ATACCTGGNAACCCCTTGGCTTGGCAGCTCTTATGGTCACTTAATTTGTGTGAAGGTC 443
Qy 223 LeuLeuAlaHisGlyAlaAspValAspSerLeuAspValIysAlaGlnThrProLeuPhe 242
Db 444 CTCCTGGAACATGGTGTCTATGTTGATAGCTTGGATGTCAAAGCACAAACACCACTTTTC 503
Qy 243 ThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSer 262
Db 504 ACTGCTGTGAGCACCGCTCATCTGGAGTCGTCGAGAATCTTTTGAAGCTGTGCTGT 563
Qy 263 ProGlyGlySerIleTyrglyHisLeuAspCysSerProValLeuThrAlaAlaArgAspGlyAla 282
Db 564 CTTAGTGTGTAGCATCTACAACTATAGCTTCTCTGCTCAGCTGCTGATGGGCT 623
Qy 283 ValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnValIysAlaIysLeu 302
Db 624 TTTGGCATCTTACAGGAGCTCTAGGCGATGTCGCGAGGCTTAATGTCAAAGCTAAACTA 683
Qy 303 ProValTrpAlaSerAsnIleAlaSerCysSerGlyProLeuTyrlleuAlaValTyrl 322
Db 684 CCAGTCTGGGCGCTTAAATATAGCTTCTGCGCCCTCTATCTGCTGCACTGCTAT 743
Qy 323 GlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrglyHis 342
Db 744 GGGCACCTTGATGTTTCCGCTGCTTTTGGCTCTATGCGGCGAGATCCCTGATTAACAATGC 803
Qy 343 ThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIleCysLeu 362
Db 804 ACTGACCGGCTTTTAAAGTGTGTTCACAGCTCGCACACTCTTGAATCTGCTT 863
Qy 363 HisAsnCysGluProGluTyrlleGlnLeuLeuIleAspPheGlyAlaAsnIleTyrl 382
Db 864 CATCATATTTGTGAGCCAGATACATCCAGCTTTTAAATAGATTTGGAGCTAACTATC 923
Qy 383 LeuProSerLeuSerLeuAspLeuThrSerGlnAspAspIysGlyIleAlaLeuLeu 402
Db 924 CTTCCATCTCTCCCTGAGCCCACTTCCAAAGATGATAAAGGCATCAAAATGCTGCTA 983
Qy 403 GlnAlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValValArgArgAla 422
Db 984 CAAGCCCGGACCTCCAGCGTCACTCTGTCCAGACCCGTTAGTTATCCGAGATCC 1043
Qy 423 LeuCysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMetLeuIle 442
Db 1044 CTCCTCGCGCGGAACCAAGTCACAAGCCAGCCAGCTGATATCCCCCTGTGTGATT 1103
Qy 443 SerTyrlleuIysHisGln 448
Db 1104 AGCTACCTCAAAACATCAA 1121

BM190670 719 bp mRNA linear EST 11-DEC-2001
LOCUS POSM0100011 E08F porcine skeletal muscle cDNA library (POSM) Sus
DEFINITION scrofa cDNA 5', mRNA sequence.
ACCESSION BM190670
VERSION 1
KEYWORDS 1 (bases 1 to 719)
SOURCE Yao, J., Cousins, P., Ernst, C. and Saama, P.
ORGANISM Sus scrofa (pig)
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
TITLE 1 (bases 1 to 719)
JOURNAL Yao, J., Cousins, P., Ernst, C. and Saama, P.
COMMENT Analysis of expressed sequence tags from a normalized porcine
skeletal muscle cDNA library
Unpublished
Contact: Jianbo Yao
```

Department of Animal Science
Michigan State University
B215 Anthony Hall, East Lansing, MI, USA
Tel: 517-355-8443
Fax: 517-353-1699
Email: yaoj@msu.edu
Seq primer: M13 reverse.
Location/Qualifiers

FEATURES

source

1..719
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/sex="male and female"
/tissue type="skeletal muscles"
/dev stage="45 d and 90 d of gestation, birth, 7 wk and 1 yr of age"
/lab_host="DH10B"
/clone_lib="porcine skeletal muscle cDNA library (POSM)"
/note="Organ: hind limbs; Vector: pSPORT1; Site_1: NotI; Site_2: SalI"

BASE COUNT 151 a 206 c 176 g 182 t 4 others
ORIGIN

Alignment Scores:

Pred. No.: 1.76e-114 Length: 719
Score: 1160.00 Matches: 222
Percent Similarity: 96.22% Conservative: 7
Best Local Similarity: 93.28% Mismatches: 8
Query Match: 49.51% Indels: 1
DB: 12 Gaps: 0

US-09-941-831A-20 (1-449) x BM190670 (1-719)

Qy 133 LeuGlnProAspIysGluGluAspThrAspThrGluGluLysGlnAlaLeuAsnGln 172
Db 7 GTCCGCGCCGCAAGGAGGAGGAGACACTGACACAGGGGAGAGCAGCCCTCAATCAA 66
Qy 173 AlaValTyrglyHisLeuAspSerTyrglyHisLeuAspGlnLeuArgGlnGluArgTyrl 192
Db 67 GCAGTGTAACACACACTCTATACCTTGGACAGCTTTTGGCCAGGAGGATACAA 126
Qy 193 ArgPheIleAsnSerArgSerGly-TripGlyValProGlyThrProLeuArgLeuAla 212
Db 127 CGATTTCATCAATAGTAGGAGTGGCTTGGGTGCTNCTTGGACACCTTGGCTGCTG 186
Qy 212 ASerTyrglyHisLeuSerCysLeuGlnValLeuLeuAlaHisGlyAlaAspValAsp 232
Db 187 TTTCTATGCCACCTTANCTGCTTGCAGTCTCTCTGCACATGGTGTGAGCTTGACAG 246
Qy 232 rleuAspValIysAlaGlnThrProLeuPheThrAlaValSerHisGlyHisLeuAspC 252
Db 247 CTTGGAGCTCAAGGACACAGACCACTTTTACCCTGCTGCTGCTGCTGCTGCTG 306
Qy 252 sValArgValLeuLeuGluAlaGlyAlaSerProGlyGlySerIleTyrglyHisLeu 272
Db 307 TGTGCTGTGCTTTTGGAGCTGGTGGCTGTCTCTGTGTGTAGCATCTACCAACTGCT 366
Qy 272 rProValLeuThrAlaAlaArgAspGlyAlaValAlaIleLeuGlnGluLeuAspHi 292
Db 367 TCCTGTGTCAAGCTGCGCTGATGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 426
Qy 292 sGlyAlaGluAlaAsnValIysAlaIysLeuProValTrpAlaSerAsnIleAlaSer 312
Db 427 TGGTCAGAGGCCAATCAAGGCAAACTACAGTCTGGTGTGCTGCTGCTGCTGCTGCT 486
Qy 312 sSerGlyProLeuTyrlleuAlaAlaValTyrglyHisLeuAspCysPheArgLeuLeu 332
Db 487 TTTGCGCCCTCTATTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 546
Qy 332 uLeuHisGlyAlaAspProAspTyrglyHisLeuAspCysThrAspGlnGlyLeuLeu 352
Db 547 GCTCCAGGAGGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 606

```

QY 352 oArgProArgThrLeuLeuGluLeuCysLeuHisAsnCysGluProGluTyrileG1 372
DB 607 ACGGCCCGACCCCTTAAATTTGCTCCACCAATAATGTGAGCCAGAGTATATCCA 666
QY 372 nLeuLeuileAspPheGlyAlaAsnileTyrLeuProSerLeuSerLeuAasp 389
DB 667 GCTGTTAATTGATTNGTGTAACTGCTACCTTNCATCTCTCCCTGGAC 718

RESULT 3
BX107270 644 bp mRNA linear EST 06-FEB-2003
LOCUS BX107270 Soares fetal heart NBH19W Homo sapiens cDNA clone
DEFINITION IMAGp998N09778 ; IMAGE:341192, mRNA sequence.
ACCESSION BX107270
VERSION BX107270.1 GI:27847318
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 644)
Ebert, L., Heil, O., Hennig, S., Neubert, P., Partsch, E., Peters, M.,
Radelof, U., Schneider, D. and Korn, B.
Human UnigeneSet - RZPD3
Unpublished
Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD; IMAGp998N09778.
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
Human UnigeneSet - RZPD3 (RZPDLIB No. 972)
http://www.rzpd.de/CloneCards/cgi-
bin/showLib.pl.cgi/response?libNo=972 Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13r, Primer sequence: TTTCACAGAAACAGCTATGAC.
FEATURES
Location/Qualifiers
1..644
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGp998N09778 ; IMAGE:341192"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal heart NBH19W"
/note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TCTTACCATCTGAGTGGAGCGCGCATCTTTTCTTTTCTTTT 3'].
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBH19W."
BASE COUNT 125 a 194 c 147 g 178 t
ORIGIN

Alignment Scores:
Pred. No.: 1,73e-109 Length: 644
Score: 113.00 Matches: 214
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 47.50% Indels: 0
DB: 13 Gaps: 0

```

```

US-09-941-831A-20 (1-449) x BX107270 (1-644)
QY 204 ProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeu 223
DB 3 CCGTGGACACCCCTTGGCTGCTCTTTATGCGCCACTTGAAGCTGTTGCAAGTCTCTC 62
QY 224 LeuAlaHisGlyAlaAaspValAaspSerLeuAaspValValAlaGlnThrProLeuPheThr 243
DB 63 TTAGCCCATGGTGGCTGTGATGTGACAGCTTGGATGCAAGGCACACGCCACTTTTCACT 122
QY 244 AlaValSerHisGlyHisLeuAaspCysValArgValLeuLeuGluAlaGlyAlaSerPro 263
DB 123 GCTGTCACTCATGGCCATCTGGACTGTGTACGTGTGCTTTTGGAGAGTGGTCCCTCTCT 182
QY 264 GlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAaspGlyAlaVal 283
DB 183 GGTGTGTAGCATCTACACAACTGTTCTCCGCTGCTCACAGCTGCCCGTGTGCTGTGT 242
QY 284 AlaIleLeuGlnLeuLeuAaspHisGlyAlaGluAlaAaspValValAlaValLeuPro 303
DB 243 GCTATCTCTGCAGGAGCTCTTAGACCATGTGTGAGAGGCCAAGCTCAAAGCTAAATACCA 302
QY 304 ValTTPAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAlaValTyrGly 323
DB 303 GTCTGGGCATCAACATAGCTTCATGTTCTGGCCCTCTTATTTGCCCGCAGCTACGGG 362
QY 324 HisLeuAaspCysPheArgLeuLeuLeuHisGlyAlaAaspProAaspTyrAsnCysThr 343
DB 363 CACCTGGAGCTGTTTCCGCTGCTTTTGTCTCCACGGGGCAGACCCCTGACTACAACTGCACT 422
QY 344 AspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluLeuCysLeuHis 363
DB 423 GACCAAGGCGCTATTGGCTGCTGCTCCAAAGCCCGCCACCTCTTGAATCTTGCCTCCAT 482
QY 364 HisAsnCysGluProGluTyrIleGlnLeuLeuIleAaspPheGlyAlaAsnIleTyrLeu 383
DB 483 CATATTTGTGAGCCAGAGTATATCCAGCTGTATATCGATTTTGGTGTCTATATCTACCTT 542
QY 384 ProSerLeuSerLeuAaspLeuThrSerGlnAaspAaspGlyIleAlaLeuLeuGln 403
DB 543 CCATCTCTCTCCCTGACCTGACCTCACAGATGATAAGGCATTCGATTCGTCTACAG 602
QY 404 AlaArgAlaThrProArgSerLeuLeuSerGlnValArgLeu 417
DB 603 GCCCGAGCCACTCCACGGTCACTTCTATCACAGGTCCTGTTA 644

RESULT 4
BX254916 861 bp mRNA linear EST 26-NOV-2002
LOCUS BU254916 603747845F1 CSEQCHN38. Gallus gallus cDNA clone CHEST655111 5', mRNA
DEFINITION sequence.
ACCESSION BU254916
VERSION BU254916.1 GI:255111095
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 861)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930

```



```

Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
  source
    1..861
      /organism="Gallus gallus"
      /mol_type="mRNA"
      /strain="White Leghorn, HiseX"
      /db_xref="taxon:9031"
      /clone="CHEST65511"
      /dev_stage="36"
      /lab_host="DH10B"
      /clone_lib="CSEQCHN38"
      /note="Organ: limbs; Vector: pBluescript II KS(+); Site 1:
      EcoRI; Site 2: NotI; This normalized library was
      constructed from 1 million independent clones. cDNA
      synthesis was initiated using an oligo(dT) primer, using
      methylated C in the first strand synthesis reaction.
      Following this first strand reaction, double-stranded cDNA
      was blunted, ligated to NotI adapters, digested with EcoRI
      , size-selected, and cloned into the NotI and EcoRI
      compatible sites of a custom modified MCS of the
      pBluescript (KS+) vector. The library was normalized in 2
      rounds using conditions adapted from Soares et al., PNAS
      (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
      (1996): 791, except that a significantly longer
      reannealing hybridization was used."
    197 a 238 c 246 g 179 t 1 others
  BASE COUNT
  ORIGIN
    Alignment Scores:
    Pred. No.: 2,77e-85 Length: 861
    Score: 890.50 Matches: 173
    Percent Similarity: 81.85% Conservative: 39
    Best Local Similarity: 66.80% Mismatches: 39
    Query Match: 38.01% Indels: 8
    DB: 13 Gaps: 3

  US-09-941-831A-20 (1-449) x BU254916 (1-861)

  QY 195 IleAsnSerArgSerGlyTrpGlyValProGlyThrProLeuArgLeuAlaSerTyr 214
  DB 6 ATCAACAGCAGGAGTGGCTGGGTGTTCACAGCAGCCCTCCACCTGCGCCCTCCAAAG 65
  QY 215 GlyHisLeuSerCysLeuGlnValLeuAlaHisGlyValAspValAspSerLeuAsp 234
  DB 66 GGCACCTCAGAGCCTGAGGTCTCTCTGTCTATGGGGCGGAGTGACAGCTGGAC 125
  QY 235 ValLysAlaGlnThrProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArg 254
  DB 126 GTGAAGGCACAAACCCCTCTCTTCACAGCAGTCAGTAATGGCCAAATGGAGTGGTGA 185
  QY 255 ValLeuLeuGluAlaGlyAlaSerProGlyGlySerIleTyrAsnAsnCysSerProVal 274
  DB 186 GCGCTCTCGAGGAGGAGTGGTCTCTGCGGAGCAGTATACAACTTGTCTCACCACGTG 245
  QY 275 LeuThrAlaAlaArgAspGlyAlaValAlaIleLeuGlnLeuLeuAspHisGlyAla 294
  DB 246 CTGACCGCAGCTAGGATGGAAACGTTGAGATTCTCAGCAGCTCTGGAAACATGTGCG 305
  QY 295 GluAlaAsnValLysAlaLysLeuProValTrpAlaSerAsnIleAlaSerCysSerGly 314
  DB 306 GAGACCAATGTCCAAGCAAGGGTGGCTGAGTGGCTGCCAATCCACTGCTGTGTCGGT 365
  QY 315 ProLeuTyrLeuAlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHis 334
  DB 366 CCGCTTTACCTCGCGGTGCTATGGGACCTGGAATGCTTTAGTGTGCTGTCTAT 425
  QY 335 GlyAlaAsnProAspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgValProArgPro 354
  DB 426 GTGGCGATCCCACTACACTCATGTGATGAGGATGATTCGGCGCATCAAGAGGCC 485
  QY 355 ArgThrLeuLeuGluIleCysLeuHisHisAsnCysGluProGluTyrIleGlnLeu 374
  DB 374
  
```

```

  DB 486 AAGACTCTGCTGGAGATCTGCTGAGGCACAGCTCCGCGGTGAGTTTCATCAAGCTGCT 545
  QY 375 IleAspPheGlyAlaAsnIleTyrLeuProSerLeuSerLeuAspLeuThrSerGlnAsp 394
  DB 546 CTGACTTTTGAGCCCAATGTGATTTGCCAAACATCAAG-----AAGATAGCAGCTGCT 599
  QY 395 AspLysGlyIleAlaLeuLeuGlnAlaArgAlaThrProArgSerLeuLeuSerGln 414
  DB 600 AGCAGAGGCGCTAGAGCTGCTGTTGCAGGCAAGAGCTCATCCCAATCCCTGATGCTCAA 659
  QY 415 ValArgLeuValValArgAlaLeuCysGlnAlaGly-----GlnProGlnAla 431
  DB 660 TCTAAGCTGGTGGTGGAGCAGCATCTCTGAAGCAGGCTGGCTGTCACATGCCCTCAG--- 716
  QY 432 IleAsnGlnLeuAspIlePro-ProMetLeuIleSerTyrLeuLysHisGlnLeu 449
  DB 717 ---AAGAGCTGGGACATTCACCCAGTCTGGTGAGCTACCTCCAAATCAGCCTT 768

  RESULT 5
  LOCUS BU220575
  DEFINITION BU220575
  ACCESSION BU220575
  VERSION BU220575.1 GI:25406909
  KEYWORDS EST.
  SOURCE Gallus gallus (chicken)
  ORGANISM Gallus gallus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
    Phasianinae; Gallus.
  REFERENCE 1 (bases 1 to 908)
  AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
    Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
  TITLE A Comprehensive Collection of Chicken cDNAs
  JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
  MEDLINE 22335534
  PUBMED 12445392
  COMMENT Contact: Simon Hubbard
    Department of Biomolecular Sciences
    University of Manchester Institute of Science and Technology (UMIST)
    PO Box 88, Manchester, M60 1QD, UK
    Tel: 01612008930
    Fax: 01612360409
    Email: Simon.Hubbard@umist.ac.uk.
  FEATURES
    source
      1..908
        /organism="Gallus gallus"
        /mol_type="mRNA"
        /strain="White Leghorn, HiseX"
        /db_xref="taxon:9031"
        /clone="CHEST4703"
        /tissue_type="whole embryo"
        /dev_stage="20-21"
        /lab_host="DH10B"
        /clone_lib="CSEQCHN04"
        /note="Organ: whole embryo; Vector: pBluescript II KS(+);
        Site 1: EcoRI; Site 2: NotI; This normalized library was
        constructed from 1 million independent clones. cDNA
        synthesis was initiated using an oligo(dT) primer, using
        methylated C in the first strand synthesis reaction.
        Following this first strand reaction, double-stranded cDNA
        was blunted, ligated to NotI adapters, digested with EcoRI
        , size-selected, and cloned into the NotI and EcoRI
        compatible sites of a custom modified MCS of the
        pBluescript (KS+) vector. The library was normalized in 2
        rounds using conditions adapted from Soares et al., PNAS
        (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
        (1996): 791, except that a significantly longer
        reannealing hybridization was used."
      223 a 231 c 244 g 210 t
  BASE COUNT
  ORIGIN
    223 a 231 c 244 g 210 t
  
```


QY 242 PheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAla 261
 |||||
 Db 226 TTCACCTGCTGACGAGCCATCGGACGTGTGACGTGCTTTTGAAGCTGTGGCC 167
 |||||
 QY 262 SerProGlyGlySerIleTyrAenAenCysSerProValLeuThrAlaAlaArgAspGly 281
 |||||
 Db 166 TGTCTGCTGTTAGCATCTACAACTGTTCTCCCGTGTCTCACAGCTGCCCGTGATGT 107
 |||||
 QY 282 AlaValAlaIleLeuGlnGluLeuAspHisGlyAlaGluAlaAsnValIlyAlaLys 301
 |||||
 Db 106 GCTGTGCTATCTGACAGAGCTCCCGAAGCGTGTGACGAGCTTTCGAAGCGT 47
 |||||
 QY 302 LeuProValTrpAlaSerAsnIleAlaSerCysSerGly 314
 |||||
 Db 46 -----AGCAATCATCAGCGGATCCGGC 23
 |||||

RESULT 7
 BU228932 849 bp mRNA linear EST 26-NOV-2002
 LOCUS 603798270F1 CSEQCHN23 Gallus gallus cDNA clone CHEST765118 5', mRNA
 DEFINITION sequence.
 ACCESSION BU228932
 VERSION BU228932.1 GI:25468283
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 849)
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Pong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 22335534
 12445392
 PUBMED

COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST
)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers

FEATURES

source
 1..849
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST765118"
 /dev_stage="22"
 /lab_host="DH108"
 /clone_lib="CSEQCHN23"
 /note="Organ: heads; Vector: pBluescript II KS(+); Site_1:
 EcoRI; Site_2: NotI; this normalized library was
 constructed from 1 million independent clones. cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was blunted, ligated to NotI adapters, digested with EcoRI
 , size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."
 198 a 224 c 241 g 186 t

Alignment Scores:

Pred. No.: 1.25e-79 Length: 849
 Score: 838.00 Matches: 158
 Percent Similarity: 84.26% Conservative: 40
 Best Local Similarity: 67.23% Mismatches: 34
 Query Match: 35.77% Indels: 3
 DB: 13 Gaps: 1
 US-09-941-831a-20 (1-449) x BU228932 (1-849)
 QY 215 GlyHisLeuSerCysLeuGlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAsp 234
 |||||
 Db 5 GGCCACCTCAGAGCGCTGGAGGTCTCTCTCTCATGGGGCGGAGGTGACAGCTGGAC 64
 |||||
 QY 235 ValIlyAlaGlnThrProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArg 254
 |||||
 Db 65 GTGAAGGCACAAACCCCTCTTTCACAGCAGTCAGTAATGGCCACTTGGAGTGGTGAAA 124
 |||||
 QY 255 ValLeuLeuGluAlaGlyAlaSerProGlyGlySerIleTyrAsnAsnCysSerProVal 274
 |||||
 Db 125 GCGCTGCTGGAGGCGAGGAGTGGTCCCTCTGGCAGCATCTACAACAATGTCTCACCCATG 184
 |||||
 QY 275 LeuThrAlaAlaArgAspGlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAla 294
 |||||
 Db 185 CTGACGGCAGCTAGGATGGGAACGTTGAGATTCTGCAGCAGCTCTGGAACATGTGTGCC 244
 |||||
 QY 295 GluAlaAsnValIlyAlaLysLeuProValTrpAlaSerAsnIleAlaSerCysSerGly 314
 |||||
 Db 245 GAGACCAATGTCACAGCAGGTGTCTGAGTGGGTGCCAATCCACTGCTGTGTCGGT 304
 |||||
 QY 315 ProLeuTyrLeuAlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHis 334
 |||||
 Db 305 CCGCTTTACCTCGCGGTGTCTAAGGCGCACCTGGAATGCTTTAGGTGTGCTGCTCTAT 364
 |||||
 QY 335 GlyAlaAspProAspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgValProArgPro 354
 |||||
 Db 365 GGTGCGCATCCCACTACCACTGATGATGAGAGGATGATTGCGCGCATCAAGGAGCCC 424
 |||||
 QY 355 ArgThrLeuLeuGluIleCysLeuHisHisAsnCysGluProGluTyrIleGlnLeuLeu 374
 |||||
 Db 425 AAGACTCTGCTGGAGATCTGCTTGAGGCACACTGCGCGGTGAGTTCATCAAGTGTCTT 484
 |||||
 QY 375 IleAspPheGlyAlaAsnIleTyrLeuProSerLeuSerLeuAspLeuThrSerGlnAsp 394
 |||||
 Db 485 CTGTGACTTTGGAGCCAAATGTGTATTGTCACCAACATCAAG-----AAGATAGCACCTG 538
 |||||
 QY 395 AspLysGlyIleAlaLeuLeuGlnAlaArgAlaThrProArgSerLeuLeuSerGln 414
 |||||
 Db 539 AGCGAGGGCTAGAGCTGCTGTTCAGGCGCAGAGCTCATCCCAATCCCTGATGTCTCAA 598
 |||||
 QY 415 ValArgLeuValValArgAlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsnGln 434
 |||||
 Db 599 TCTAGGCTGTGTGAGGCACATCTCTGAAGCAGGCTGGCTGTGCACATGCCCTCAGAGAG 658
 |||||
 QY 435 LeuAspIleProPro-MetLeuIleSerTyrLeuLysHisGln 448
 |||||
 Db 659 CTGGACATTCACCACTGCTGCTGAGTACCTCCCAACATCAG 701
 |||||
 RESULT 8
 CA364256 713 bp mRNA linear EST 06-NOV-2002
 LOCUS 639132 NCCOWA 1RT Oncorhynchus mykiss cDNA clone 1RT86F15_B_C08 5',
 DEFINITION mRNA sequence.
 ACCESSION CA364256
 VERSION CA364256.1 GI:24673292
 KEYWORDS EST.
 SOURCE Oncorhynchus mykiss (rainbow trout)
 ORGANISM Oncorhynchus mykiss
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
 1 (bases 1 to 713)
 Rextroad, C.E. and Keele, J.W.
 Sequence analysis of a rainbow trout normalized cDNA library

BASE COUNT
 ORIGIN

Unpublished
Contact: Rexroad CE
USDA, ARS, National Center for Cool and Cold Water Aquaculture
11876 Leetown Road, Kearneysville, WV 25430, USA
Tel: 304 724 8340 x2129
Fax: 304 725 0351
Email: crexroad@nccwa.ars.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified by
cross_match v0.990329.
Seq primer: AGCGGATACAAATTTTCACACAGGA.

```

FEATURES
source
Location/Qualifiers
i. .713
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/clone="1RT86F15_B_C08"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="NCCCWA 1RT"
/note="Vector: pCMV SPOT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from brain, gill, liver,
spleen, muscle, and kidney."
BASE COUNT 165 a 193 c 215 g 140 t
ORIGIN
Alignment Scores:
Pred. No.: 3.02e-76
Score: 805.50
Percent Similarity: 81.94%
Best Local Similarity: 66.52%
Query Match: 34.38%
DB: 14
Matches: 151
Conservative: 35
Mismatches: 40
Indels: 1
Gaps: 1

```

US-09-941-831A-20 (1-449) x CA364256 (1-713)

140	Qy	LysMetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGlu	159
30	Db	AAAGTGAAGTCTAAATGCAGCTCTCCCAAGATCTTCTCTGTTCTCAGCCC--	86
160	Qy	GluAspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSer	179
87	Db	GATGAGGAGGCAATAGCATGTGGCCAGGCTCTGAACCAGGACGAGTGAATGATG	146
180	Qy	TyrThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSer	199
147	Db	GAGGTGTTAGCTGAGCTGTTTCCCGAGGAGCTACAGAAAGTCTATCAACACAG	206
200	Qy	GlyTyrGlyValProGlyThrProLeuArgLeuAlaSerTyrGlyHisLeuSerCys	219
207	Db	GGCTGGGGGATCCCTGTTTACCCCTCTACGGACTCCCGCAGCACACGAGACCT	266
220	Qy	LeuGlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuLeuAspValLysAla	239
267	Db	CTGGAGCTCTCTGCTGGACACGGAGCGGAGGTGGACAGTCTAGATGTGAGG	326
240	Qy	ProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGlu	259
327	Db	CTCTGTTTTACAGCGGTCACTGGTTAAACACTGGACTGTGTTGTGGTCTTACT	386
260	Qy	GlyAlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAla	279
387	Db	GGAGCCGACCCCAACGGCGAGCCGCTACAAACAACCTGTTCCCGGTGTCAGC	446
280	Qy	AspGlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnVal	299
447	Db	GAGGAGACGTGGAGGTCTCTCAGGAGCTGCTTCAGTTCGGAGCCGAGGTGCG	506
300	Qy	AlaLysLeuProValTyrAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAla	319
507	Db	CCCAAGTCCCTGAGTGGGCCCTCCACGCCACAGCTGCAGGGGACCCCTGTAC	566
320	Qy	AlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAsp	339

[illegible]FEATURES
source

```

seq primer: 17:
Location/Qualifiers
1..758
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/clone="tcaa0002c.o.21"
/tissue_type="adipose tissue, blood, brain,
differentiating gonads, gills, interrenal, intestine,
kidney, liver, muscle, ovary, pituitary, testis"
/dev_stages="from embryos to adults"
/lab_host="DH10B"
/clone_lib="AGENAS Rainbow trout multi-tissues library
(tcaa)"
/note="Vector: p773D-pac; Clone distribution : AGENA
Resource centre: Francois.PIUMI,
Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie e
Etude du genome (LRGE), Domaine de Vilvert, 78352,
Jouy-en-Josas cedex, France"
202 a 177 c 176 q 202 t 1 others
BASE COUNT

```

BASE COUNT	202 a	177 c	176 q	202 t	1 others
Jouy-en-Josas cedex, France"					

Alignment Scores:		
Pred. No.:	3.33e-76	Length:
Score:	805.50	Matches:
Percent Similarity:	79.51%	Conservative:
Best Local Similarity:	61.89%	Mismatches:
Query Match:	34.38%	Indels:
DB:	13	Gaps:
		1
		1
		758

US-09-941-831A-20 (1-449) x BX073379 (1-758)

Qy 142 AsnLeuMetAspIleThrLysIle---PheSerLeuLeuGlnProAspLysGluGluGlu 160

BASE COUNT 125 a 130 c 134 g 129 t
ORIGIN

Alignment Scores:

Pred. No.: 1,07e-70 Length: 518
Score: 752.00 Matches: 148
Percent Similarity: 91.86% Conservatives: 10
Best Local Similarity: 86.05% Mismatches: 14
Query Match: 32.10% Indels: 1
DB: 28 Gaps: 0

US-09-941-831A-20 (1-449) x AZ725513 (1-518)

QY 152 LeuLeuGlnProAspLysGluGluAspThrGluGluLysGlnAlaLeuAsn 171
DB 2 CTCTGCAACCTGAAGAGGAGGAGGACACTGGCACCAGGGAACAGGCTCTGAAT 61
QY 172 GlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeuLeuArgGlnGluArgTyr 191
DB 62 CAAGCTGTCTATGACAACTGACTGTGACCTCGGACAC-CTTCTACACAGGAACGTTAT 120
QY 192 LysArgPheIleAsnSerArgSerGlyTyrGlyValProGlyThrProLeuArgLeuAla 211
DB 121 AAACGGCTCATCAACAGCACGAGTGGCTGGGGGATTCCTGGAACACCCCTTGCGAATGGCA 180
QY 212 AlaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAlaHisGlyAlaAspValAsp 231
DB 181 GCTTCCATGGACACTTAATGTGTGAAGGGCTCTCTGAACATGCTGTGATGTTGAT 240
QY 232 SerLeuAspValLysAlaGlnThrProLeuPheThrAlaValSerHisGlyHisLeuAsp 251
DB 241 AGCTTGAGTGTCAAGACACAAACACACATTTTCACTGCTGTGAGCCACCGTCTATCTGGAG 300
QY 252 CysValArgValLeuLeuGluAlaGlyAlaSerProGlyGlySerIleTyrAsnAsnCys 271
DB 301 TGGCTGAGAATGCTTTTAGAAGCTGGTGCCTGTCTAGTGGTAGCATCTACAAATTCG 360
QY 272 SerProValLeuThrAlaAlaArgAspGlyAlaValAlaIleLeuGlnLeuLeuAsp 291
DB 361 TCTCTGTTCTCACTGCCTCAGGTGATGGGGCTTTTGCCATCTTACAGAGCTCCTAGGG 420
QY 292 HisGlyAlaGluAlaAsnValLysAlaLysLeuProValTyrAlaSerAsnIleAlaSer 311
DB 421 CATGTTGCCGAGGCTAATGTCAAGCTAACTACCACTAGTCTGGGCGTCAATATAGCTTCA 480
QY 312 CysSerGlyProLeuTyrLeuAlaAlaValTyrGly 323
DB 481 TGTCTGCCCCCTCTATCTGCTGCTGAGTCTATGGG 516

RESULT 13

LOCUS CA358977 680 bp mRNA linear EST 05-NOV-2002
DEFINITION 631831 NCCOWA 1RT Oncorhynchus mykiss cDNA clone 1RT5D15_B_08 5', mRNA sequence.

ACCESSION CA358977 GI:24604164

VERSION CA358977.1

KEYWORDS EST.

SOURCE Oncorhynchus mykiss (rainbow trout)

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

1 (bases 1 to 680)

REFERENCE Rextroad,C.E. and Keele,J.W.

AUTHORS Sequence analysis of a rainbow trout normalized cDNA library

TITLE Unpublished

JOURNAL

COMMENT Contact: Rextroad CF

USDA, ARS, National Center for Cool and Cold Water Aquaculture

11876 Leetown Road, Kearneysville, WV 25430, USA

Tel: 304 724 8340 x2129

Fax: 304 725 0351

Email: crexroad@nccowa.ars.usda.gov

Single pass sequencing. Bases called with phred v0.020425.c and

trimmed with the aid of the trim_alt option. Vector identified by
cross_match v0.990329.

Seq primer: AGCGGATAACAATTTTCACACAGGA.

FEATURES

source location/Qualifiers
1..680
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/clone="1RT5D15_B_08"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="NCCOWA 1RT"
/note="Vector: PCWV SPORT6; Site 1: NotI; Site 2: SalI;
library made from pooled tissue from brain, gill, liver,
spleen, muscle, and kidney." 159 a 185 c 196 g 140 t

BASE COUNT 159 a 185 c 196 g 140 t
ORIGIN

Alignment Scores:

Pred. No.: 3.69e-69 Length: 680
Score: 739.50 Matches: 141
Percent Similarity: 82.76% Conservatives: 27
Best Local Similarity: 69.48% Mismatches: 34
Query Match: 31.56% Indels: 1
DB: 14 Gaps: 1

US-09-941-831A-20 (1-449) x CA358977 (1-680)

QY 140 LysMetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGluGlu 159
DB 69 AAGATGAGTCTAATGATGTCTCAAGATCTTCTCTCTCTTCAGCCC---AAGGAAGAG 125
QY 160 GluAspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAspSer 179
DB 126 GATGAGGAGCAATGAGCATGCGCAGGCTCTGAACGAGGAGTGTAGAGTGTATGACGTA 185
QY 180 TyrThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArgSer 199
DB 186 GAGGTCTTAGCTGAGCTGTGTCTCCAGGAGAGCTACAGAAAGTCTATCAACAGCAGAGC 245
QY 200 GlyTyrGlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCys 219
DB 246 GGTCTGGGGATCCCTGTATTACCCCTCTACGAGCTGCCGAGCAGCACCGACACCTGAGGTGT 305
QY 220 LeuGlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThr 239
DB 306 CTGGAGCTCTCTGGAGCAGGAGCGGAGGTGGACAGTCTAGATGTGAAGGCCAGACC 365
QY 240 ProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAla 259
DB 366 CTTCTGTTTACAGCGGTCAAGTGTGTAACACCTGGAGCTGTGTGCTCTTACTGAGGCT 425
QY 260 GlyAlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArg 279
DB 426 CGAGCCGACCCCAACCGCAGCCCGTACAACTGTTCCTCCGCTGCTGACCGCCGCGCGC 485
QY 280 AspGlyAlaValAlaIleLeuGlnLeuLeuAspHisGlyAlaGluAlaAsnValLys 299
DB 486 GAGGGAGCGTGGAGGTCTCTCAGGGAGCTGTTCAGTTCGGAGCCGAGGTCTGACGCTGG 545
QY 300 AlaLysLeuProValTyrAlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAla 319
DB 546 CCCAAGTCTCTGAGTGGGCTTCCACGCCACAGCCTGCGAGGAGCCCTGTACATATACC 605
QY 320 AlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspProAsp 339
DB 606 GCTGTATATGGACACCTGGACTGTTTTTAAGTGTCTGTCTCCACGGGGCTAATCTCAAC 665
QY 340 TyrAsnCys 342
DB 666 TATAACTGT 674

RESULT 14

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 1, 2003, 08:00:27 ; Search time 510 Seconds
(without alignments)
9940.284 Million cell updates/sec

Title: US-09-941-831A-6

Perfect score: 1878

Sequence: 1 catgattacgcgaagcttgg.....taaaaaaaaaaaaaaaaaa 1878

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq 19Jun03:*

- 1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
- 25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1878	100.0	1878	22	AAH46857 Human serine/threose
2	1878	100.0	1878	22	AAH78730 Human HIBCJ89 seri
3	1878	100.0	1878	22	AAH33372 DNA encoding human
C 4	1877.2	100.0	2530	22	AAH33364 DNA encoding human
5	1117.8	59.5	1222	24	ABK11107 Human polynucleoti
6	989.8	52.7	1127	24	ABZ11533 Human polynucleoti
7	931	49.6	1069	24	ABK11106 DNA encoding human
C 8	709.8	37.8	8214	23	AAH77257 DNA encoding novel

9	703.4	37.5	1345	24	ABN60012 Novel human coding
10	241.8	12.9	2353	22	AAF63581 Human phosphatase
11	181.4	9.7	3047	23	ABL20003 Drosophila melanog
12	179.8	9.6	2786	23	ABL06295 Drosophila melanog
C 13	169	9.0	1828	22	ABA19208 Human nervous syst
C 14	169	9.0	1828	22	AAK70743 Human immune/haema
C 15	149.6	8.0	3169	23	ABL17297 Drosophila melanog
C 16	141	7.5	408	23	ABV61004 Human prostate exp
17	135.8	7.2	2482	24	ABQ93336 Human cDNA SQ ID
18	108.8	5.8	2019	19	AAV38672 Mus musculus SOCS7
19	103.6	5.5	3382	24	ABK83507 Human cDNA differe
C 20	100.8	5.4	1266	22	AAS33071 DNA encoding human
C 21	100.8	5.4	1433	23	ABV23231 Human prostate exp
22	100.8	5.4	1433	23	ABV29074 Human prostate exp
23	100.8	5.4	1509	25	AAD49617 Human cytoskeleton
24	100.8	5.4	1540	22	AAS33289 DNA encoding human
25	91	4.8	545	23	AAS73111 DNA encoding novel
C 26	85.4	4.5	220	22	ABA13268 Human nervous syst
27	85.4	4.5	2575	23	AAS73882 DNA encoding novel
28	72.8	3.9	1966	22	AAH14752 Human cDNA sequenc
29	72.8	3.9	2710	22	AAK52009 Human polynucleoti
30	72.8	3.9	4330	24	AAD40741 Human kinase and p
31	71.4	3.8	550	21	AAC94977 Cat flea hindgut a
32	71.2	3.8	389	24	ABL82288 Homo sapiens SOCS7
33	71.2	3.8	419	19	AAV38673 Human polynucleoti
34	71.2	3.8	2346	22	AAK52993 Human cDNA clone (
35	70	3.7	802	22	AAH07276 Human ovarian anti
36	69.6	3.7	1024	24	ABQ54306 cDNA encoding huma
37	69.4	3.7	401	21	AAH77749 Colon tumour relat
38	69.4	3.7	401	22	AAI28487 Human colon tumour
39	69.4	3.7	401	25	ABZ32673 Human tyrosine pho
40	67.2	3.6	3411	18	AAI91998 CviJI coding seque
41	66.6	3.5	5496	15	AAQ73395 Drosophila melanog
C 42	65.8	3.5	8397	23	ABL20002 Drosophila melanog
C 43	65.8	3.5	9648	23	ABL06294 Drosophila melanog
C 44	65.8	3.5	9648	23	ABL17296 Human OREF ORF1001
45	64.4	3.4	1434	21	AAC75446

ALIGNMENTS

RESULT 1

AAH46857
ID AAH46857 standard; cDNA; 1878 BP.

XX AC AAH46857;

XX DT 25-SEP-2001 (first entry)

XX DE Human serine/threonine phosphatase encoding cDNA (clone ID HIBCJ89).

XX KW Serine/threonine phosphatase; nootropic; neuroprotective; cytoskeletal; dermatological; immunosuppressive; antiinflammatory; antibacterial; ss;
XX KW anti-HIV; antiparkinsonian; antisticking; antianemic; antiarthritic;
XX KW antirheumatic; virucide; hepatotropic; cerebrotropic; vulnerary;
XX KW antiinflammatory; nephrotropic; gene therapy; vaccine.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
XX FT CDS 367..1716
XX FT /*tag= a

XX PN WO200155388-A1.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01395.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 02-MAR-2000; 2000US-0186350.

XX (HUMA-) HUMAN GENOME SCI INC.
XX PA
XX PI Rosen CA, Barash SC, Ruben SM;
XX DR WPI; 2001-476208/51.
XX DR P-PSDB; AAB85477.
XX DR
XX PT Novel proteins of serine/threonine phosphatase family, useful for
XX PT diagnosing, treating, preventing and/or prognosing disorders related to
XX PT the proteins, including cancer, immune response and neuronal disorders
XX PT
XX PS Claim 1; Page 426; 441pp; English.
XX CC The invention provides serine/threonine phosphatase family polypeptides
XX CC and polynucleotides encoding them. The polypeptides can be expressed by
XX CC standard recombinant methodology. The polypeptides, polynucleotides and
XX CC antibodies are useful for diagnosis, prognosis, prevention and treatment
XX CC of neurodegenerative disorders, immune system disorders, autoimmune
XX CC diseases, allergic reactions, infectious diseases, hyperproliferative
XX CC disorders, renal disorders, cardiovascular disorders, cerebrovascular
XX CC disorders, respiratory disorders, endocrine disorders, gastrointestinal
XX CC disorders and also muscular, reproductive disorders (see AAB46846 for a
XX CC detailed description of the diseases that can be treated). They are also
XX CC useful as immune system enhancers, immunosuppressive agents, stimulator
XX CC of B-cell responsiveness to pathogens, activator of T-cells, to induce
XX CC higher affinity antibodies, as an agent to increase serum immunoglobulin
XX CC concentrations, to accelerate recovery of immunocompromised individuals,
XX CC to boost immunoresponsiveness in aged populations and/or neonates, as
XX CC regulator of antigen presentation, as a means to induce tumour
XX CC proliferation in pathologies e.g. AIDS, and/or common variable immuno
XX CC deficiency. The polypeptides and polynucleotides are useful to prevent
XX CC skin aging, for preventing hair loss, to stimulate growth and
XX CC differentiation of hemopoietic cells and bone marrow cells, for
XX CC supporting cell culture of primary tissues, to modulate mammalian
XX CC characteristics such as body weight, height, eye color, hair color and
XX CC skin, to modulate mammalian metabolism to change a mammal's mental or
XX CC physical state, and as food additive or preservative. The present
XX CC sequence represents a human serine/threonine phosphatase encoding cDNA.
XX SQ Sequence 1878 BP; 479 A; 499 C; 417 G; 483 T; 0 other;

Query Match 100.0%; Score 1878; DB 22; Length 1878;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATGATTACGCCAAGCTTGGCAGCGGGTAGAAAACGTGAAATTTGGTGGATCGTTATGTG 60
DB |||||||
DB 1 CATGATTACGCCAAGCTTGGCAGCGGGTAGAAAACGTGAAATTTGGTGGATCGTTATGTG 60
QY 61 AGTAAGAAACCGACTAATGGATTTCTTTATCTTACTGCAACCAACCTGATCTATGTGGAG 120
DB |||||||
DB 61 AGTAAGAAACCGACTAATGGATTTCTTTATCTTACTGCAACCAACCTGATCTATGTGGAG 120
QY 121 GCTTCAGGTGCAGCCCGGAAAGAAACATGGATGCACTCCATCACAATTGCCACTGTGGAG 180
DB |||||||
DB 121 GCTTCAGGTGCAGCCCGGAAAGAAACATGGATGCACTCCATCACAATTGCCACTGTGGAG 180
QY 181 AAGTTACCATCACTAGCTGGGTCTCCCTGACCTCCGCTGCAAGAAATTTCCGGGTG 240
DB |||||||
DB 181 AAGTTACCATCACTAGCTGGGTCTCCCTGACCTCCGCTGCAAGAAATTTCCGGGTG 240
QY 241 GCCCACTTTGTTTAGATTTCTGACCTTGTGTGTCATGAGTTTATATTTCACTGCTCAAG 300
DB |||||||
DB 241 GCCCACTTTGTTTAGATTTCTGACCTTGTGTGTCATGAGTTTATATTTCACTGCTCAAG 300
QY 301 CTTTCTCAGCCAGCATTAACAGATCTTTATGCTTTTCTTATATCCCAATCCTCA 360
DB |||||||
DB 301 CTTTCTCAGCCAGCATTAACAGATCTTTATGCTTTTCTTATATCCCAATCCTCA 360
QY 361 AAAGAGATGAGGGAAGTGGATGGAACCTGATTGACCCCAATATCAAGCTTTGGCGGTATG 420
DB |||||||

QY 1501 GCTAATATCTACCTTCCATCTCTCCCTTGACCTGACCTCACAAGATGATAAAGCAT 1560
 DB |||||
 QY 1501 GCTAATATCTACCTTCCATCTCTCCCTTGACCTGACCTCACAAGATGATAAAGCAT 1560
 DB |||||
 QY 1561 GCATTGCTGCTACAGCCGCGACCACTCCACGGTCACTTCTATCATCAGGTCCGTTAGTC 1620
 DB |||||
 QY 1561 GCATTGCTGCTACAGCCGCGACCACTCCACGGTCACTTCTATCATCAGGTCCGTTAGTC 1620
 DB |||||
 QY 1621 GTCGCCAGAGCTGTGCGCAGCTGCGCCAGCCACAGCCATCAACAGCTGATATTCCT 1680
 DB |||||
 QY 1621 GTCGCCAGAGCTGTGCGCAGCTGCGCCAGCCACAGCCATCAACAGCTGATATTCCT 1680
 DB |||||
 QY 1681 CCCATGTTGATTAGCTACCTTAAACACCACTGTATCTTGCAGTCTCCCCAGGAACTTA 1740
 DB |||||
 QY 1681 CCCATGTTGATTAGCTACCTTAAACACCACTGTATCTTGCAGTCTCCCCAGGAACTTA 1740
 DB |||||
 QY 1741 TGATGCTCCGAAACCACTCGGGGACCTCAGTAGCTGAGAGCATTAACAGCTCATCCA 1800
 DB |||||
 QY 1741 TGATGCTCCGAAACCACTCGGGGACCTCAGTAGCTGAGAGCATTAACAGCTCATCCA 1800
 DB |||||
 QY 1801 CTTACTGAGCTGCTCTCTCTGTTATCTCTCCACATTAATTTCTCCAGAAATAAGTA 1860
 DB |||||
 QY 1801 CTTACTGAGCTGCTCTCTCTGTTATCTCTCCACATTAATTTCTCCAGAAATAAGTA 1860
 DB |||||
 QY 1861 AAAAAAAAAAAAAAAAAA 1878
 DB |||||
 QY 1861 AAAAAAAAAAAAAAAAAA 1878
 DB |||||

RESULT 2

AAH78730
 ID AAH78730 standard; cDNA; 1878 BP.
 AC AAH78730;
 DT 03-DEC-2001 (first entry)
 DE Human HIBCU89 serine/threonine phosphatase cDNA sequence.
 KW Human; HIBCU89; ss; serine/threonine phosphatase; PSpase; vaccine;
 KW gene therapy; PSpase expression; PSpase expression; PSpase modulation;
 KW immune disorder; autoimmune disorder; Wiscott-Aldrich syndrome;
 KW Chediak-Higashi syndrome; Hashimoto's thyroiditis; multiple sclerosis;
 KW inflammation; Crohn's disease; inflammatory bowel disease; appendicitis;
 KW rheumatoid arthritis; cellular proliferative disorder; lymphoma;
 KW lung cancer; intestinal cancer; cardiovascular disorder; aneurysm;
 KW Scimitar syndrome; Ebstein's anomaly.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 CDS 367..1716
 FT /*tag= a
 FT /product= "HIBCU89 PSpase protein"
 XX
 PN WO200164703-A1.
 XX
 PD 07-SEP-2001.
 XX
 PF 28-FEB-2001; 2001WO-US06256.
 XX
 PR 02-MAR-2000; 2000US-0186350.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ebner R, Ruben SM;
 XX
 DR WPI; 2001-530113/58.
 DR P-PSDB; AAG77803.
 XX
 PT Nucleic acids encoding serine/threonine phosphatase polypeptides,
 useful for preventing, diagnosing and/or treating, e.g. Crohn's

disease, lung cancer and Scimitar syndrome -
 Claim 1; Page 308; 335pp; English.

The present sequence represents the specifically claimed human cDNA clone HIBCU89 which shares homology with members of the serine/threonine family of phosphatases. The invention comprises novel human serine/threonine phosphatase (PSpase) polypeptides and polynucleotides which may be used in the prevention (vaccine), diagnosis and treatment (gene therapy) of diseases associated with inappropriate PSpase expression. The PSpase polynucleotides of the invention may be used as DNA probes to detect and quantitate the presence of similar nucleic acids in samples. The PSpase polypeptides may be used as antigens in the production of antibodies against the PSpase polypeptides and in assays to identify modulators of PSpase expression and activity. The anti-PSpase antibodies and antagonists may also be used to down regulate expression and activity, the anti-PSpase antibodies may also be used as diagnostic agents for detecting the presence of PSpase polypeptides in samples. Disorders that may be prevented, diagnosed and/or treated by the invention are: immune/autoimmune disorders (e.g. Wiscott-Aldrich syndrome, Chediak-Higashi syndrome, Hashimoto's thyroiditis and multiple sclerosis); inflammatory conditions (e.g. Crohn's disease, inflammatory bowel disease, appendicitis and rheumatoid arthritis); cellular proliferative disorders (e.g. lymphoma, lung and intestinal cancers); and cardiovascular disorders (e.g. Scimitar syndrome, Ebstein's anomaly and aneurysm).

Sequence 1878 BP; 479 A; 499 C; 417 G; 483 T; 0 other;

Query Match 100.0%; Score 1878; DB 22; Length 1878;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CATGATTACGCCAAGCTTGGCAGGAGGTAGAAAACGTGAATTTGGTGGATCGTTATGTG 60
 DB |||||
 QY 1 CATGATTACGCCAAGCTTGGCAGGAGGTAGAAAACGTGAATTTGGTGGATCGTTATGTG 60
 DB |||||
 QY 61 AGTAAGAAACACAGCTAATGGGATCTTTATCTTACTGCAACCCACCTGATCTATGTGGAG 120
 DB |||||
 QY 61 AGTAAGAAACACAGCTAATGGGATCTTTATCTTACTGCAACCCACCTGATCTATGTGGAG 120
 DB |||||
 QY 121 GCTTCAGGTGTCAGCCGCGAAAGAAACATGGATTGCACATCCATCATTGCCACTGTGGAG 180
 DB |||||
 QY 121 GCTTCAGGTGTCAGCCGCGAAAGAAACATGGATTGCACATCCATCATTGCCACTGTGGAG 180
 DB |||||
 QY 181 AAGTTACCATCAGTAGCTGGGTTGTCCTGACCTCCGCTGCAAGAAATTCGGGTG 240
 DB |||||
 QY 181 AAGTTACCATCAGTAGCTGGGTTGTCCTGACCTCCGCTGCAAGAAATTCGGGTG 240
 DB |||||
 QY 241 GCCCACTTTGTTTATAGATTCTGACCTGTGTGCCATGAGTTTATATTTTCACTGCTCAAG 300
 DB |||||
 QY 241 GCCCACTTTGTTTATAGATTCTGACCTGTGTGCCATGAGTTTATATTTTCACTGCTCAAG 300
 DB |||||
 QY 301 CTTTCTCAGCCAGCATTACCTGAAGATCTTTATGCTTTTTTCTTATAATCCCAATCCTCA 360
 DB |||||
 QY 301 CTTTCTCAGCCAGCATTACCTGAAGATCTTTATGCTTTTTTCTTATAATCCCAATCCTCA 360
 DB |||||
 QY 361 AAAGAGATGAGGAAAGTGGATGGAACCTGATTGACCCCAATATCAGACTTTGGGCGTATG 420
 DB |||||
 QY 361 AAAGAGATGAGGAAAGTGGATGGAACCTGATTGACCCCAATATCAGACTTTGGGCGTATG 420
 DB |||||
 QY 421 GGAATACCCACAGAAACTGGACCATACAGATGCCAGAAACTATCAGATATGCAGC 480
 DB |||||
 QY 421 GGAATACCCACAGAAACTGGACCATACAGATGCCAGAAACTATCAGATATGCAGC 480
 DB |||||
 QY 481 ACCTACCTCTCTGAAATAGTGGTTCTTAAATCTGTTTACCTGGGAACGGTGGTGAAGT 540
 DB |||||
 QY 481 ACCTACCTCTCTGAAATAGTGGTTCTTAAATCTGTTTACCTGGGAACGGTGGTGAAGT 540
 DB |||||
 QY 541 TCAAAAGTTGAGAAAGTAAAGAACGTGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
 DB |||||
 QY 541 TCAAAAGTTGAGAAAGTAAAGAACGTGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
 DB |||||

CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
 CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
 CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
 CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
 CC angina and thrombosis), infections caused by bacteria, viruses and
 CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
 CC agonists, antagonists and antibodies can also be used to promote wound
 CC healing, maintain organs before transplantation, and support cell culture
 CC of primary tissues. AAS3043-AAS3486 represent human secreted protein
 CC coding sequences, PCR primers, and related sequences of the invention.
 CC Note: the sequence data for this patent did not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO
 CC at: ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 1878 BP; 479 A; 499 C; 417 G; 483 T; 0 other;

Query Match 100.0%; Score 1878; DB 22; Length 1878;

Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CATGATTACGCAAGCTTGGCAGAGGGTAGAAAACGTGAATTTGGTGATCGTTATG	60
DB	1	CATGATTACGCAAGCTTGGCAGAGGGTAGAAAACGTGAATTTGGTGATCGTTATG	60
QY	61	AGTAAGAACCCAGCTTAATGGGATCTTTATCTTACTGCAACCCACCTGATCTATGG	120
DB	61	AGTAAGAACCCAGCTTAATGGGATCTTTATCTTACTGCAACCCACCTGATCTATGG	120
QY	121	GCTTCAGGTGACGCGGAGAAACATAGGATGACCTCCATCACAATTCGCACTGTGG	180
DB	121	GCTTCAGGTGACGCGGAGAAACATAGGATGACCTCCATCACAATTCGCACTGTGG	180
QY	181	AAGTTACCATCATATGACCTGGGTTGTCCTGACCTCCGCTGCAAGAAATTCGGGG	240
DB	181	AAGTTACCATCATATGACCTGGGTTGTCCTGACCTCCGCTGCAAGAAATTCGGGG	240
QY	241	GCCTCTGTTTATGATCTGACCTGTTGTCGATGAGGTTTATATTTCACTGCTCAAG	300
DB	241	GCCTCTGTTTATGATCTGACCTGTTGTCGATGAGGTTTATATTTCACTGCTCAAG	300
QY	301	CTTTCTCAGCAGCATTAAGATCTTTATGCTTTTCTTATTAATCCCAATCCTCA	360
DB	301	CTTTCTCAGCAGCATTAAGATCTTTATGCTTTTCTTATTAATCCCAATCCTCA	360
QY	361	AAAGAGATGAGGAAAGTGGATGAAACTGATTGACCAATATCAGACTTTGGCGTATG	420
DB	361	AAAGAGATGAGGAAAGTGGATGAAACTGATTGACCAATATCAGACTTTGGCGTATG	420
QY	421	GGATATCCCAACAGAACTGGACCAATAACAGATGCCAAGAACTATGAGATATGCAG	480
DB	421	GGATATCCCAACAGAACTGGACCAATAACAGATGCCAAGAACTATGAGATATGCAG	480
QY	481	ACCTACCTCTCTGAAATAGTGGTTCTCTAAATCTGTTACCTTGGGAACGGTGGAA	540
DB	481	ACCTACCTCTCTGAAATAGTGGTTCTCTAAATCTGTTACCTTGGGAACGGTGGAA	540
QY	541	TCAAAGTTGAGAGTAAAGAACGTGTCCTGCTCTCTGGAATTTTACACTCGCTGTG	600
DB	541	TCAAAGTTGAGAGTAAAGAACGTGTCCTGCTCTCTGGAATTTTACACTCGCTGTG	600
QY	601	GCTGCCATTTGGCGCTGTAGCCAGCCTCTCTCTGGATTTTACCTCGCTGTGATGAT	660
DB	601	GCTGCCATTTGGCGCTGTAGCCAGCCTCTCTCTGGATTTTACCTCGCTGTGATGAT	660
QY	661	GAGCTCTTGTGGAGCCATTAGCCAAAACAAACCCAGGGAGCCAGTTTATGATGTTGA	720
DB	661	GAGCTCTTGTGGAGCCATTAGCCAAAACAAACCCAGGGAGCCAGTTTATGATGTTGA	720
QY	721	GACACAGACCAAGATCTGGCAATTTCTTGTGCTCATATAGAGATAGTTCTTCAATTA	780
DB	721	GACACAGACCAAGATCTGGCAATTTCTTGTGCTCATATAGAGATAGTTCTTCAATTA	780

QY	781	GCCAGATGAACCTCATGGAATCACCAGATCTTCTCCCTCTCGAGCCCGACAAGGAG	840
DB	781	GCCAGATGAACCTCATGGAATCACCAGATCTTCTCCCTCTCGAGCCCGACAAGGAG	840
QY	841	GAGGAGGACATGACACAGAGAGAGAGGCTCTCAATCAAGCAGTGTATGACAAACGAC	900
DB	841	GAGGAGGACATGACACAGAGAGAGAGGCTCTCAATCAAGCAGTGTATGACAAACGAC	900
QY	901	TCCTATACCTTTGGACCACTTTGGCCAGAGCGGTACAAACGTTTCATCAACAGCAGG	960
DB	901	TCCTATACCTTTGGACCACTTTGGCCAGAGCGGTACAAACGTTTCATCAACAGCAGG	960
QY	961	AGTGGCTGGGCTGTTCTTGGGACACCTTGGCGCTGGCTTCTTATGGCCACTTTGAC	1020
DB	961	AGTGGCTGGGCTGTTCTTGGGACACCTTGGCGCTGGCTTCTTATGGCCACTTTGAC	1020
QY	1021	TGTTTGAAGTCTCTTAGCCCATGGTGTGATGTTGACAGCTTGGATGTCAAGGCACAG	1080
DB	1021	TGTTTGAAGTCTCTTAGCCCATGGTGTGATGTTGACAGCTTGGATGTCAAGGCACAG	1080
QY	1081	AGCCACCTTTTCACTGCTCAGTCAATGCGCATCTGGACTGTGACGTGCTTTTGGAA	1140
DB	1081	AGCCACCTTTTCACTGCTCAGTCAATGCGCATCTGGACTGTGACGTGCTTTTGGAA	1140
QY	1141	GCTGTGCTCTCTCTGGTGTAGCATCTACAACAACTGTTCTCCCGTGTCTACAGCTGC	1200
DB	1141	GCTGTGCTCTCTCTGGTGTAGCATCTACAACAACTGTTCTCCCGTGTCTACAGCTGC	1200
QY	1201	CGTGTGCTGCTGTTGCTTCTCTGAGGAGTCTCTAGACCATGTTGTCAGAGGCCAACGTC	1260
DB	1201	CGTGTGCTGCTGTTGCTTCTCTGAGGAGTCTCTAGACCATGTTGTCAGAGGCCAACGTC	1260
QY	1261	AAAGCTAACTACAGTCTGGGCATCAACATAGCTTCTGTTCTGGCCCCCTCTATTG	1320
DB	1261	AAAGCTAACTACAGTCTGGGCATCAACATAGCTTCTGTTCTGGCCCCCTCTATTG	1320
QY	1321	GCCTGAGTCTACGGGCACCTGGACCTGTTTCCGCTGCTTTTCTCCAGGGGACAGCCCT	1380
DB	1321	GCCTGAGTCTACGGGCACCTGGACCTGTTTCCGCTGCTTTTCTCCAGGGGACAGCCCT	1380
QY	1381	GACTACAACTGACCTGACCGGCTTATGGCTCGTGTCTCCAAAGACCCCGCACCTCTT	1440
DB	1381	GACTACAACTGACCTGACCGGCTTATGGCTCGTGTCTCCAAAGACCCCGCACCTCTT	1440
QY	1441	GAAATCTGCTCCATCAATAATTTGTGAGCCAGAGTATATCCAGCTGTTAATCGATTTG	1500
DB	1441	GAAATCTGCTCCATCAATAATTTGTGAGCCAGAGTATATCCAGCTGTTAATCGATTTG	1500
QY	1501	GCTAATATCTACCTTCCATCTCTCTCTCTGACCTGACCTCAACAGATGATAAGGCATT	1560
DB	1501	GCTAATATCTACCTTCCATCTCTCTCTCTGACCTGACCTCAACAGATGATAAGGCATT	1560
QY	1561	GCAATGCTGTACAGGCGGAGCCACTCCAGGTCACCTTCTATACAGGTCGGTTAGTC	1620
DB	1561	GCAATGCTGTACAGGCGGAGCCACTCCAGGTCACCTTCTATACAGGTCGGTTAGTC	1620
QY	1621	GTCCGAGAGCCTTGTGCGAGGCTGGCGACCAAGGCCATCAACAGCTGATATTCCT	1680
DB	1621	GTCCGAGAGCCTTGTGCGAGGCTGGCGACCAAGGCCATCAACAGCTGATATTCCT	1680
QY	1681	CCCATGTTGATTTAGCTTACCTAAACCAACCTGTAATCTTGAGTCTCTCCAGGAATTA	1740
DB	1681	CCCATGTTGATTTAGCTTACCTAAACCAACCTGTAATCTTGAGTCTCTCCAGGAATTA	1740
QY	1741	TGATGCTCTCCGAAAAACCACTGGGACCTCACGTAGCTGGAGAGCAATTCAGGCTCATCCA	1800
DB	1741	TGATGCTCTCCGAAAAACCACTGGGACCTCACGTAGCTGGAGAGCAATTCAGGCTCATCCA	1800
QY	1801	CTTACCTGAGGCTGCTCTCTGATTTATCTCTCAAAATAAATTTCTCCAGAAAAATAAGTA	1860
DB	1801	CTTACCTGAGGCTGCTCTCTGATTTATCTCTCAAAATAAATTTCTCCAGAAAAATAAGTA	1860

Db 1041 TCCTATATTTGGACACAGCTTTTGGCCGAGGAGCGTTACAAAGCTTTTCATCAACAGCAGG 982
QY 961 AGTGGCTGGGGTGTCTCCCTGGGACACCTTTGGCGTCTGCTTCTTATGCGCACATTGAGC 1020
Db 981 AGTGGCTGGGGTGTCTCCCTGGGACACCTTTGGCGTCTGCTTCTTATGCGCACATTGAGC 922
QY 1021 TGTTCGCAAGTCTCTTAGCCCATGGTGTGTGTTGACAGCTTGGATGTCGAAGGCACAG 1080
Db 921 TGTTCGCAAGTCTCTTAGCCCATGGTGTGTGTTGACAGCTTGGATGTCGAAGGCACAG 862
QY 1081 AGCCCACTTTTCACTGCTGTCACTGATGATGAGCCATCTGGACTGTGTAGCTGTGCTTTGGAA 1140
Db 861 AGCCCACTTTTCACTGCTGTCACTGATGATGAGCCATCTGGACTGTGTAGCTGTGCTTTGGAA 802
QY 1141 GCTGGTGCTCTCTCGTGTAGCTATCAACAACCTGTTCTCCCGTGTCTACAGCTGCC 1200
Db 801 GCTGGTGCTCTCTCGTGTAGCTATCAACAACCTGTTCTCCCGTGTCTACAGCTGCC 742
QY 1201 CGTGATGGTGTCTGCTATCTCTGAGGAGCTCTTAGACCATGGTGCAGAGCCCAACGTC 1260
Db 741 CGTGATGGTGTCTGCTATCTCTGAGGAGCTCTTAGACCATGGTGCAGAGCCCAACGTC 682
QY 1261 AAAGCTAACTACAGCTCTGGGATCAACAATAGCTTCAATGTTCTGGCCCCCTCTATTGG 1320
Db 681 AAAGCTAACTACAGCTCTGGGATCAACAATAGCTTCAATGTTCTGGCCCCCTCTATTGG 622
QY 1321 GCCGAGCTCTACGGGACCTGGACTGTTTCCGCTCTGTTTCCGACGGGCGAGACCT 1380
Db 621 GCCGAGCTCTACGGGACCTGGACTGTTTCCGCTCTGTTTCCGACGGGCGAGACCT 562
QY 1381 GACTCAACTGCACTGACGAGGCGCTTATGGCTGTGTCCTCAAGACCCCGCACCTCTCTT 1440
Db 561 GACTCAACTGCACTGACGAGGCGCTTATGGCTGTGTCCTCAAGACCCCGCACCTCTCTT 502
QY 1441 GAAATCTGCTCCATCATATTTGAGCCAGAGTATATCAGCTGTATTCAGATTTGTT 1500
Db 501 GAAATCTGCTCCATCATATTTGAGCCAGAGTATATCAGCTGTATTCAGATTTGTT 442
QY 1501 GCTAATATCTACCTTCCTCTCTCTGCTGACCTGACCTCAAGATGATTAAGGCATT 1560
Db 441 GCTAATATCTACCTTCCTCTCTCTGCTGACCTGACCTCAAGATGATTAAGGCATT 382
QY 1561 GCATTGCTGTACAGGCGGAGCCACTCCACGGTCACTTCTATCAAGCTCGTGTAGTC 1620
Db 381 GCATTGCTGTACAGGCGGAGCCACTCCACGGTCACTTCTATCAAGCTCGTGTAGTC 322
QY 1621 GTCGCGAGGCTTTGTCGAGGCTGCGCCAGCCCAAGCCATCAACAGCTGGATATTCCT 1680
Db 321 GTCGCGAGGCTTTGTCGAGGCTGCGCCAGCCCAAGCCATCAACAGCTGGATATTCCT 262
QY 1681 CCCATCTGATTTAGCTACCTTAACACCACTGTTATCTTGCAGTCTCCCGAGGACTTA 1740
Db 261 CCCATCTGATTTAGCTACCTTAACACCACTGTTATCTTGCAGTCTCCCGAGGACTTA 202
QY 1741 TGATGCTCTCGAAACACCTCGGGGACTCACGTAGCTGGAGACATTAACAGCTCTATCCA 1800
Db 201 TGATGCTCTCGAAACACCTCGGGGACTCACGTAGCTGGAGACATTAACAGCTCTATCCA 142
QY 1801 CTTACTGGAGCTGCTCTCTCTGTTATTATCTTCCACAAATAAAATTTCTCCAGAAATAAGTA 1860
Db 141 CTTACTGGAGCTGCTCTCTCTGTTATTATCTTCCACAAATAAAATTTCTCCAGAAATAAGTA 82
QY 1861 AAAAAAAAAAAAAAAAAA 1878
Db 81 AAAAAAAAAAAAAAAAAA 64

RESULT 5

ABK1107

ID ABK11107 standard; DNA; 1222 BP.

XX

AC

ABK11107;

XX

DT 05-JUN-2002 (first entry)
XX DNA encoding human NOV5b protein, homologue of ankyrin-repeat proteins.
DE Human; NOVX-associated disorder; developmental disorder; blood disorder;
XX endocrine disorder; vascular disease; gastrointestinal disorder; cancer;
KW respiratory disorder; inflammatory disorder; reproductive disorder;
KW neurodegenerative disorder; autoimmune disorder; infectious disease;
KW cardiovascular disorder; cell signal processing; ankyrin-repeat protein;
KW metabolic pathway modulation; NOV5b; gene; ds.
XX Homo sapiens.
OS
FH Key Location/Qualifiers
FT CDS 6..1088
FT /*tag= a
FT /*product= "NOV5b"
XX
PN WO200206329-A2.
PD 24-JAN-2002.
XX
PF 18-JUL-2001; 2001WO-US22709.
XX
PR 18-JUL-2000; 2000US-218870P.
PR 18-JUL-2000; 2000US-218875P.
PR 18-JUL-2000; 2000US-218901P.
PR 24-JUL-2000; 2000US-220273P.
PR 26-JUL-2000; 2000US-220912P.
PR 27-JUL-2000; 2000US-221233P.
PR 28-JUL-2000; 2000US-221650P.
XX (CURA-) CUPAGEN CORP.
XX
PI Rastelli L, Shinkets RA, Zerhusen B, Malyankar UM, Padigaru M;
XX WPI; 2002-179781/23.
DR P-PSDB; AAU77411.
XX
PT Novel cytoplasmic, nuclear, membrane bound and secreted NOVX
PT polypeptides, useful for treating developmental disorders, endocrine
PT disorders, vascular disorders, infectious diseases and
PT neurodegenerative disorders -
XX
PS Claim 9; Page 39; 178pp; English.
XX
CC The present invention relates to the isolation of novel human
CC polypeptides referred to as NOV1, NOV2, NOV3, NOV4a, NOV4b, NOV5a,
CC NOV5b, NOV6 AND NOV7, and the polynucleotide sequences encoding them.
CC The NOVX polypeptides are related to NOPE, cadherin, interferon
CC alpha-13, ADAM, ankyrin repeat-containing, transpanin or semaphorin
CC polypeptides. The sequences of the invention are useful for identifying
CC an agent (a cellular receptor or downstream effector) that binds to
CC the NOVX polypeptide, or an agent that modulates it's expression or
CC activity. They are useful for treating or preventing NOVX-associated
CC disorders such as developmental disorders, endocrine disorders, vascular
CC diseases, gastrointestinal disorders, reproductive disorders, inflammatory
CC disorders, blood disorders, autoimmune and immune disorders, infectious diseases,
CC cardiovascular disorders, cancers, and other disorders related to cell
CC signal processing and metabolic pathway modulation. The present sequence
CC encodes the human NOV5b protein.
XX
SQ Sequence 1222 BP; 276 A; 352 C; 278 G; 316 T; 0 other;

Query Match 59.5%; Score 1117.8; DB 24; Length 1222;

Best Local Similarity 99.8%; Pred. No. 2.2e-306;

Matches 1119; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 730 CCAAGATCTGCATTCTCTGTGCTCATTAATGAGATAGTTCTCAATAGCCAGATG 789

DB 102 CCACAGATCTGCATTCTCTGTGCTCATTAATGAGATAGTTCTCAATAGCCAGATG 161

```
QY 790 AACCTCATGACATCAACCAAGATCTTCTCCTCTCGAGCCCGACAGGAGGAGGAC 849
Db 162 AACCTCATGACATCAACCAAGATCTTCTCCTCTCGAGCCCGACAGGAGGAGGAC 221
QY 850 ACTGACACAGAGAGAGGAGGCTCTCAATCAAGCAGTGTATGACAAAGCTCTCTACT 909
Db 222 ACTGACACAGAGAGAGGAGGCTCTCAATCAAGCAGTGTATGACAAAGCTCTCTACT 281
QY 910 TTGGACAGCTTTTGGCCAGGAGCGTTACAAAGCTTTATCAACAGCAGGAGTGTGG 969
Db 282 TTGGACAGCTTTTGGCCAGGAGCGTTACAAAGCTTTATCAACAGCAGGAGTGTGG 341
QY 970 GGTGTTCTGGGACACCTTGGCCCTGGCTCTTCTTATGGCCACTTGGCTTTGCAA 1029
Db 342 GGTGTTCTGGGACACCTTGGCCCTGGCTCTTCTTATGGCCACTTGGCTTTGCAA 401
QY 1030 GTCTCTTAGCCCATGGTGTGATGTGACAGCTTGGATGTCAGGACAGAGCCACTT 1089
Db 402 GTCTCTTAGCCCATGGTGTGATGTGACAGCTTGGATGTCAGGACAGAGCCACTT 461
QY 1090 TTCACTGCTGTAGTATGAGCCATCTGGACTGTGACGTGTCTTTTGAAGCTGTGCG 1149
Db 462 TTCACTGCTGTAGTATGAGCCATCTGGACTGTGACGTGTCTTTTGAAGCTGTGCG 521
QY 1150 TCTCTGTGTAGTACTTCAACAACTGTCTCCCGTGTCTACAGCTGCCCTGTATGT 1209
Db 522 TCTCTGTGTAGTACTTCAACAACTGTCTCCCGTGTCTACAGCTGCCCTGTATGT 581
QY 1210 GCTGTTGCTATCTGACAGAGCTCTTAGACCATGGTGCAGAGCCCAAGCTCAAGCTAA 1269
Db 582 GCTGTTGCTATCTGACAGAGCTCTTAGACCATGGTGCAGAGCCCAAGCTCAAGCTAA 641
QY 1270 CTACCACTCTGGGCATCAACATAGCTTCAATGTTCTGGCCCTCTTATTTGGCCGAGTC 1329
Db 642 CTACCACTCTGGGCATCAACATAGCTTCAATGTTCTGGCCCTCTTATTTGGCCGAGTC 701
QY 1330 TAGGGCAGCTGGAGCTGTTTTCGCCCTGCTTTCCTCCAGGGGACAGCCCTGACTACAAC 1389
Db 702 TAGGGCAGCTGGAGCTGTTTTCGCCCTGCTTTCCTCCAGGGGACAGCCCTGACTACAAC 761
QY 1390 TGCACTGACAGGCGCTATTTGGCTGTGTCGACAGCCCGACCTCTTGAATCTGC 1449
Db 762 TGCACTGACAGGCGCTATTTGGCTGTGTCGACAGCCCGACCTCTTGAATCTGC 821
QY 1450 CTCATCATATTTGACGACAGATATATCCAGCTGTTAATCGATTTTGGTGTATATC 1509
Db 822 CTCATCATATTTGACGACAGATATATCCAGCTGTTAATCGATTTTGGTGTATATC 881
QY 1510 TACCTTCCATCTCTCTCCCTTGACCTGACCTCAAGAGATGATAAGGCAATTGCATTGCTG 1569
Db 882 TACCTTCCATCTCTCTCCCTTGACCTGACCTCAAGAGATGATAAGGCAATTGCATTGCTG 941
QY 1570 CTACAGGCGGAGCCACTCAAGCTCACTTCTATCAAGCTCGTTTGTGTCGTCGACAG 1629
Db 942 CTACAGGCGGAGCCACTCAAGCTCACTTCTATCAAGCTCGTTTGTGTCGTCGTCGACAG 1689
QY 1630 GCCTTGTGCGAGCTGCGGACGACAGCCATCAACAGCTGATATCTCTCCAGTTG 1689
Db 1002 GCCTTGTGCGAGCTGCGGACGACAGCCATCAACAGCTGATATCTCTCCAGTTG 1061
QY 1690 ATTAGCTACCTAAACACCACTGATATCTTGCAGTCTCCCGAGGAACCTTATGATCCCTC 1749
Db 1062 ATTAGCTACCTAAACACCACTGATATCTTGCAGTCTCCCGAGGAACCTTATGATCCCTC 1121
QY 1750 CGAAACACCTGGGAGCTCAAGTGTGAGAGCATTTACGCTCATCACTTACCTGG 1809
Db 1122 CGAAACACCTGGGAGCTCAAGTGTGAGAGCATTTACGCTCATCACTTACCTGG 1181
QY 1810 AGCTGCTCTCTGTTATTTCTCCCAATAAATTTCTCCAG 1850
Db 1182 AGCTGCTCTCTGTTATTTCTCCCAATAAATTTCTCCAG 1222
```

RESULT 6

```
ABZ11533
ID ABZ11533 standard; cDNA; 1127 BP.
XX ABZ11533;
AC ABZ11533;
XX
DT 20-JAN-2003 (first entry)
XX
XX Human polynucleotide SEQ ID NO 415.
DE
XX
XX Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW cell-proliferative disorder; neurodegenerative disease; bacterial;
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW arthritis; cytostatic; immunomodulator; neotropic; neuroprotective;
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW haemostatic; vulnary; fungicide; antibacterial; virucide; protozoacide;
KW antiarthritic; gene; ss.
XX
OS Homo sapiens.
XX
XX WO200270539-A2.
PN
XX
PD 12-SEP-2002.
XX
XX 05-MAR-2002; 2002WO-US05095.
PF
XX
XX 05-MAR-2001; 2001US-0799451.
PR
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren P;
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
PI Wehrman T, Wang J, Wang D, Drmanac RT;
XX
XX WPI; 2002-759812/82.
DR
XX
XX P-PSDB; ABP69316.
XX
XX New polynucleotides comprising sequences assembled from expressed
PT sequence tags (ESTs), useful for treating cell-proliferative,
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
PT platelet or coagulation disorders
XX
XX Claim 1: SEQ ID NO 415; 1012pp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated polynucleotide (I) comprising a
CC nucleotide sequence selected from any of 948 sequences
CC (ABZ1119-ABZ12066) or their mature protein coding portion, active domain
CC coding protein or complementary sequences. The polynucleotides are useful
CC for identifying expressed genes or for physical mapping of human genome.
CC The encoded polypeptides (ABP6902-ABP6949) are useful as molecular
CC weight markers, as a food supplement, for generating antibodies, in
CC medical imaging, screening and diagnostic assays and for treating
CC cell-proliferative disorders (cancer), neurodegenerative diseases
CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
CC disorders, platelet or coagulation disorders, wound, burns, incision,
CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
CC parasitic), arthritis, etc.
CC
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1127 BP; 244 A; 329 C; 281 G; 273 T; 0 other;
SQ
```

Query Match 52.7%; Score 989.8; DB 24; Length 1127;
Best Local Similarity 99.8%; Pred. No. 4.5e-270;
Matches 991; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 735 GATCTGGCATTTCTTGTGCTCATATGAGATAGTTCTCCATAGGCAAGTGAACCT 794
Db 134 GATCTGGCATTTCTTGTGCTCATATGAGATAGTTCTCCATAGGCAAGTGAACCT 193

[illegible]

RESULT 9	
ABN60012	
ID	ABN60012 standard; cDNA; 1345 BP.
XX	
XX	AC
ABN60012;	
XX	
DT	28-JUN-2002 (first entry)
XX	
DE	Novel human coding sequence SEQ ID NO: 423.
XX	
KW	Human; antianaemic; vulnuary; antiinflammatory; immunomodulator;
KW	antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW	neuroprotective; antiparkinsonian; protein therapy; Est;
KW	expressed sequence tag; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200222660-A2.
XX	
PD	21-MAR-2002.
XX	
PF	10-SEP-2001; 2001WO-US26015.
XX	
PR	11-SEP-2000; 2000US-0659671.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI	Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX	
DR	WPI; 2002-292408/33.
DR	P-PSDB; ABB97599.
XX	
PT	An isolated polynucleotide for treating diseases associated with its
PT	encoded polypeptide such as cancer and multiple sclerosis -
XX	
PS	Claim 1; SEQ ID NO 423; 509pp; English.
XX	
CC	The present invention provides the protein and coding sequences of 444
CC	novel human proteins. These were isolated from expressed sequences tags
CC	(ESTs). They can be used to stimulate cell growth, to regulate

CC	haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC	e.g. in burn treatment, to regulate the immune system e.g. to treat
CC	multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC	infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC	stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC	e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC	Parkinson's disease. The present sequence is a coding sequence of the
CC	invention.
XX	
SQ	Sequence 1345 BP; 380 A; 300 C; 321 G; 344 T; 0 other;
	Query Match 37.5%; Score 703.4; DB 24; Length 1345;
	Best Local Similarity 99.2%; Pred. No. 9e-189; 6; Indels 0; Gaps 0;
	Matches 707; Conservative 0; Mismatches
QY	23 CGAGGGTAGAAAACGTGAAATTGGTGGATCGTTATGTGAGTAAGAAACACAGCTAATGGGA 82
DB	111 CCNAGGTAGAAAACGTGAAATTGGTGGATCGTTATGTGAGTAAGAAACACAGCTAATGGGA 170
QY	83 TTCTTTTATCTTACTGCAACCCACCTGATCTATGTGGAGCTTTCAGGTGAGCCGCCGAAAG 142
DB	171 TTCTTTTATCTTACTGCAACCCACCTGATCTATGTGGAGCTTTCAGGTGAGCCGCCGAAAG 230
QY	143 AAACATGGATTGCACCTCCATCACATTCGCCACTGTGAGAGATTACCCATCCTAGCCTGG 202
DB	231 AAACATGGATTGCACCTCCATCACATTCGCCACTGTGGAGATTACCCATCCTAGCCTGG 290
QY	203 GTTGTCCTCCCTGACCCCTCCGCTGCAAGAAATTCGGGGTGGCCACTTTGTTTGTAGATTCTG 262
DB	291 GTTGTCCCTGACCCCTCCGCTGCAAGAAATTCGGGGTGGCCACTTTGTTTGTAGATTCTG 350
QY	263 ACCTTGTGTGCCATGAGGTTTATTTCACTGCTCAAGCTTTCTCAGCCAGCATTAACCTG 322
DB	351 ACCTTGTGTGCCATGAGGTTTATTTCACTGCTCAAGCTTTCTCAGCCAGCATTAACCTG 410
QY	323 AAGATCTTTATGCTTTTCTTATTAATCCCAATCCTCAAAAGAGATGAGGGAAGTGGAT 382
DB	411 AAGATCTTTATGCTTTTCTTATAATCCCAATCCTCAAAAGAGATGAGGGAAGTGGAT 470
QY	383 GGAACCTGATTGACCCCAATATCAGACTTTTGGGGCTATGGGAATACCCACAGAAACTGGA 442
DB	471 GGAACCTGATTGACCCCAATATCAGACTTTTGGGGCTATGGGAATACCCACAGAAACTGGA 530
QY	443 CCATAACAGATGCCAACAGAAACTATGAGATATGCAGCACCTACCCCTCCTGAAATAGTGG 502
DB	531 CCATAACAGATGCCAACAGAAACTATGAGATATGCAGCACCTACCCCTCCTGAAATAGTGG 590
QY	503 TTCTTAAATCTGTTACTTTGGGAAACGGTGGTTGGAAGTTCAAAGTTCAGAGTTAAAGAAC 562
DB	591 TTCTTAAATCTGTTACTTTGGGAAACGGTGGTTGGAAGTTCAAAGTTCAGAGTTAAAGAAC 650
QY	563 GTGTCCTGTGTCCTCCTACTCTCAAAAGAGAAACAATGCTGCATTTGCGCGCTGTAGCC 622
DB	651 GTGTCCTGTGTCCTCCTACTCTCAAAAGAGAAACAATGCTGCATTTGCGCGCTGTAGCC 710
QY	623 AGCCTCTCTGGAATTTTACACTCGCTGTGTAGATGATGAGCTCTTTGTTGGAGGCCATT 682
DB	711 AGCCTCTCTGGAATTTTACACTCGCTGTGTAGATGATGAGCTCTTTGTTGGAGGCCATT 770
QY	683 GCCAAACAAACCCAGGGAGCCAGTTTATGTATGTTGTAGACACAGACCAAAG 735
DB	771 GCCAAACAAACCCAGGGAGCCAGTTTATGTATGTTGTAGACACAGACCAAAG 823
RESULT 10	
AAF63581	
ID	AAF63581 standard; cDNA; 2353 BP.
XX	
AC	AAF63581;
XX	
XX	11-MAY-2001 (first entry)
DT	
DE	Human phosphatase MTMR7 h coding sequence.

XX	Phosphatase; gene therapy; enzyme; cancer; pathophysiological hypoxia;
KW	cardiac dysfunction; vascular disorder; myopathy; ectodermal dysplasia;
KW	congenital muscle disorder; Papillon-Lefevre syndrome; Cowden disease;
KW	Moebius syndrome; Bjornstad syndrome; Bannayan Zonana syndrome;
KW	schizophrenia; hamartoma; ss.
XX	
OS	homo sapiens.
XX	
XX	WO200112819-A2.
XX	
PD	22-FEB-2001.
XX	
PF	11-AUG-2000; 2000WO-US222158.
XX	
PR	13-AUG-1999; 99US-0149005.
XX	
PA	(SUGB-) SUGEN INC.
XX	
PI	Plowman GD, Martinez R, Whyte D, Hill R, Flanagan P, Lioubin M;
XX	
DR	WPI: 2001-211226/21.
DR	P-PSDB; AAB73229.
XX	
PT	New protein phosphatase polypeptide for diagnosing and treating
PT	phosphatase related disorders such as cancer, schizophrenia, cardiac
PT	dysfunction and/or vascular disorders
XX	
PS	Claim 1; Fig 4; 138pp; English.
XX	
CC	The present invention relates to phosphatase proteins and coding
CC	sequences. The present sequence is one such phosphatase coding sequence.
CC	Phosphatases are enzymes that catalyse the dephosphorylation of proteins
CC	modified by phosphorylation of serine, threonine or tyrosine residues.
CC	The phosphatases are useful for treating a variety of diseases: for
CC	example cancer e.g. breast, urogenital, prostate, head, neck, lung
CC	cancers, synovial sarcomas, renal cell carcinoma, non-small cell lung
CC	cancer, hepatocellular carcinoma, pancreatic endocrine tumours, stomach
CC	cancer, glioblastoma, colorectal cancer and thyroid cancer.
CC	pathophysiological hypoxia, cardiac dysfunction and/or vascular
CC	disorders, myopathies, congenital muscle disorders, Papillon-Lefevre
CC	syndrome, Cowden disease, ectodermal dysplasia, Moebius syndrome,
CC	Bjornstad syndrome, Bannayan Zonana syndrome, schizophrenia and
CC	hamartomas.
XX	
SQ	Sequence 2353 BP; 670 A; 506 C; 535 G; 642 T; 0 other;
	Query Match 12.9%; Score 241.8; DB 22; Length 2353;
	Best Local Similarity 58.3%; Pred. No. 1.2e-57;
	Matches 423; Conservative 0; Mismatches 302; Indels 0; Gaps 0;
QY	10 GCCAGCTTGGCACAGGGTAGAAACGTGAATTTGGTGGATCGTTATGTGAGTACGAA 69
Db	239 GCTCAGGGAAGCCGAAAGGTTGAAATGTCGCTTGGTAGTCGAGTGTCTCTCAAAAA 298
QY	70 CCAGCTAATGGGATTCTTTATCTTACTGCAACCCACCTGATCTATGTGGAGGCTTCAGT 129
Db	299 GCAGCTCTAGGTACTTTGTTATTTGACGCTACCATGTCTATTCGTGGAATAATTCACCT 358
QY	130 GCAGCCCCGGAAGAAACATGGATTGCACTCCATCCATTCGCCACTGTGGAGAAAGTTACCC 189
Db	359 GACGCAAGAAAGAAACATGGATTCTTCACAGTCAGATTTCACCATTTGAGAAACAGGCA 418
QY	190 ATCACTAGCTGGTGTGTCCTCGTACCCTCGCTGCAAGAAATTCGGGTGGCCCACTTT 249
Db	419 ACAACCGCTACCGGATGCCCTCTGCTGATTCGCTGCAAGAACTTTTCAGATAATACAGCTC 478
QY	250 GTTTTAGATTCTGACCTTGTGTGTCATCAGGCTTTTATTTTCACTGTCTCAAGCTTTCTCAG 309
Db	479 ATCATACCTCAGGAAGAGATTGCCACGAGTGTACATCTCCCTTGATACGCCTTGCAAG 538
QY	310 CCAGCATTAACCTGAAGATCTTTATGCTTTTTTTTATTAATCCCAAAATCCTCAAAAGAGATG 369

Db 634 GACGCTGTGCTCCATGAACGAGAGTAGACAGCTGTGCGACACCTATCCGCGTCAGATTTA 693
QY 501 GGTTCCTAAATCTGTTACCTTTGGGAACGGTGGTTGGAAGTTCAAAGTTCAAGAGTAAGA 560
Db 694 CGTGCCCAAGGAGCCACACGCTGATGCTCATCAGCAGCTCGGATTCGCTCCAAAGG 753
QY 561 ACCTGTCCTGCTGCTCTCTACCTCTTACAAGAGAACAAATGCTGCCATTTGCCGCTGTAG 620
Db 754 GCGGCTGCCAGTGTCTACCTATCTGCACAAC---AACAAAGCTTCATCTGCCGCTGCAG 810
QY 621 CCAGCCTCTCTCTGGATTTTACACTCGCTGTAGATGAGCTCTTTGTTGGAGCCAT 680
Db 811 TCAGCCCTCTCCGGATTCTAGTCCCGCTCTCTGGAGATGAGCAGATGCTGGAGGCCAT 870
QY 681 TAGCAAAACAAACCCAGGGAGCCAGTTTATGTATGTTGTAGACACACCAAGACCAAGAT 737
Db 871 ACGCAGACGAAATCCNACAGGACTACATGATGTGTTGGACACCGGACCCGGAT 927

RESULT 13
ABAI9208/c
ID ABAI9208 standard; DNA; 10828 BP.
XX ABAI9208;
AC ABAI9208;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 11539.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antithumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antituber; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180828.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.

14-AUG-2000; 2000US-0225758.
14-AUG-2000; 2000US-0225759.
18-AUG-2000; 2000US-0226279.
22-AUG-2000; 2000US-0226681.
22-AUG-2000; 2000US-0226868.
22-AUG-2000; 2000US-0227182.
23-AUG-2000; 2000US-0227009.
30-AUG-2000; 2000US-0228924.
01-SEP-2000; 2000US-0229287.
01-SEP-2000; 2000US-0229343.
01-SEP-2000; 2000US-0229344.
01-SEP-2000; 2000US-0229345.
05-SEP-2000; 2000US-0229509.
05-SEP-2000; 2000US-0229513.
06-SEP-2000; 2000US-0230437.
06-SEP-2000; 2000US-0230438.
08-SEP-2000; 2000US-0231242.
08-SEP-2000; 2000US-0231243.
08-SEP-2000; 2000US-0231244.
08-SEP-2000; 2000US-0231413.
08-SEP-2000; 2000US-0231414.
08-SEP-2000; 2000US-0232080.
08-SEP-2000; 2000US-0232081.
12-SEP-2000; 2000US-0231968.
14-SEP-2000; 2000US-0232397.
14-SEP-2000; 2000US-0232398.
14-SEP-2000; 2000US-0232399.
14-SEP-2000; 2000US-0232400.
14-SEP-2000; 2000US-0232401.
14-SEP-2000; 2000US-0233063.
14-SEP-2000; 2000US-0233064.
14-SEP-2000; 2000US-0233065.
21-SEP-2000; 2000US-0234223.
21-SEP-2000; 2000US-0234274.
25-SEP-2000; 2000US-0234997.
25-SEP-2000; 2000US-0234998.
26-SEP-2000; 2000US-0235484.
27-SEP-2000; 2000US-0235834.
27-SEP-2000; 2000US-0235836.
29-SEP-2000; 2000US-0236327.
29-SEP-2000; 2000US-0236367.
29-SEP-2000; 2000US-0236368.
29-SEP-2000; 2000US-0236369.
29-SEP-2000; 2000US-0236370.
02-OCT-2000; 2000US-0236802.
02-OCT-2000; 2000US-0237037.
02-OCT-2000; 2000US-0237038.
02-OCT-2000; 2000US-0237039.
02-OCT-2000; 2000US-0237040.
13-OCT-2000; 2000US-0239935.
13-OCT-2000; 2000US-0239937.
20-OCT-2000; 2000US-0240960.
20-OCT-2000; 2000US-0241785.
20-OCT-2000; 2000US-0241786.
20-OCT-2000; 2000US-0241787.
20-OCT-2000; 2000US-0241808.
20-OCT-2000; 2000US-0241809.
20-OCT-2000; 2000US-0241826.
20-OCT-2000; 2000US-0242221.
01-NOV-2000; 2000US-0244617.
08-NOV-2000; 2000US-0246474.
08-NOV-2000; 2000US-0246475.
08-NOV-2000; 2000US-0246476.
08-NOV-2000; 2000US-0246477.
08-NOV-2000; 2000US-0246478.
08-NOV-2000; 2000US-0246523.
08-NOV-2000; 2000US-0246524.
08-NOV-2000; 2000US-0246525.
08-NOV-2000; 2000US-0246526.
08-NOV-2000; 2000US-0246527.
08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246609.


```
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 03-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUNA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metatasis -
XX
XX Disclosure; SEQ ID NO 2555; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 10828 BP; 3640 A; 2147 C; 1989 G; 3052 T; 0 other;
SQ
Query Match 9.0%; Score 169; DB 22; Length 10828;
Best Local Similarity 92.2%; Pred. No. 1.2e-36;
Matches 178; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 282 TTATATTTCACTGCTCAAGCTTTCTCAGCCAGCATTACTGAGATCTTTATGCTTTTTC 341
Db 6241 TTTTGTTCATGCTATCTCTCTCTCCAGCATTACTGAGATCTTTATGCTTTTTC 6182
QY 342 TTATAATCCAAATCTCTCAAGAGATGAGGGAAGTGGATGGAACCTGATTGCCCAAT 401
Db 6181 TTATAATCCAAATCTCTCAAGAGATGAGGGAAGTGGATGGAACCTGATTGCCCAAT 6122
```

QY 402 ATCAGACTTTGGCGGTATGGGATATCCCAACAGAAACTGGACCATACAGATGCCAACAG 461
DB 6121 ATCAGACTTTGGCGGTATGGGATATCCCAACAGAAACTGGACCATACAGATGCCAACAG 6062
QY 462 AAACATATGAGATTA 474
DB 6061 AAACATATGAGGTA 6049

RESULT 15

ABL17297
ID ABL17297 standard; DNA; 3169 BP.
XX AC ABL17297;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 3364.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX Claim 1; SEQ ID NO 3364; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 3169 BP; 816 A; 877 C; 796 G; 680 T; 0 other;

Query Match 8.0%; Score 149.6; DB 23; Length 3169;
Best Local Similarity 54.6%; Pred. No. 2e-31;
Matches 321; Conservative 0; Mismatches 264; Indels 3; Gaps 1;

QY 150 GATTGACATCCATCATTGCGCTGAGAGAGTTACCCATCATTAGCTGGGTTGTCC 209
DB 726 GATCTCGACATGCGCTGCGGAGTTTCCCTTGAGCAGCAGGATCTCC 785
QY 210 CTGACCTCCGCTGCAAGATTTCCGGTGGCCACTTGTGTTAGATCTGACCTTGT 269
DB 786 GCTACTCATCCGCTGCAAGACCTTCTCTCCGTAACCTTTCGTCATTTCCCAAGGACTCCGA 845
QY 270 GTGCCATGAGGTTTATATTTCACTGCTCAAGCTTTCTCAGCCAGCATTTACCTGAAGATCT 329

DB 846 GTGCCACGATGCTACACTTCGCTGCTGAAACTCTTCCAGCCGGTGTCCATCAACAAAT 905
QY 330 TTATGCTTTTCTTATAATCCCAATCTCRAAAGAGATAGGGAAAGTGGATGAAACT 389
DB 906 GTACTGCTTTCAACTACCGCGGAGTTCAGGATGATTTCCCAAGAACGCTGGCTGGGATTA 965
QY 390 GATTGACCCCAATPATCAGACTTTTGGGCGTATGGGAATACCCAAACAGAAACTGGACCATTAAC 449
DB 966 CTTTAAACTGGAGCGGAGTTTCAAGCACATGCTGGTGCCCAACGAGGCCCTGGACGCTGTG 1025
QY 450 AGATGCCAACAGAAACTATGAGATATGACAGACCTACCTCTCTGAAATAGTGTTCCTAA 509
DB 1026 CTCCATGAACGAGAAAGTACGAGCTGTGCGACACCTATCCGCGTCAAGTTTACGTGCCCA 1085
QY 510 ATCTGTTACCTTTGGGAACGCTGTTTGGAGTTTCAAAAGTTTCAAGAGTAAAGAACGCTGCC 569
DB 1086 GGAGGCCACCGCTGATGCTCATCAGCAGCTCGGATTTCCGCTCCNAGGGGGGCTGCC 1145
QY 570 TGTGCTCTCTACCTCTTACAAAGAGAAACAATGCTGCCATTTTCGCTGTAGCCAGCTCT 629
DB 1146 AGTGTCTACCTATCTGCACAAC--AACAAGGCTTCCATCTGCGCTGCAGTCAGCCCT 1202
QY 630 CTCTGATTTTACACTCGCTGTGTAGATGATGAGCTCTTGTGGAGGCCATTTAGCCAAAC 589
DB 1203 GTCCGGATTCAGTGCCCGCTGTCTGGAGGATGAGCAGATGCTGGAGGCCATACGCAAGAC 1262
QY 690 AAACCCAGGAGCCAGTTTATGCTATGTTGTAGACACACAGACCAAGAT 737
DB 1263 GAATTTCAACCGGACTACATGATGTTGTGAGACACGCGCGGAT 1310

Search completed: December 1, 2003, 10:38:15

Job time : 516 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2003, 07:34:53 ; Search time 39 Seconds
(without alignments)
541.410 Million cell updates/sec

Title: US-09-941-831A-20
Perfect score: 2343
Sequence: 1 MRESGWKLIDPISDFGRMG.....QAINQLDIPPLMISYLKQL 449

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1599	68.2	308	1 AS12_HUMAN	Q8wxk4 homo sapien
2	499	21.3	335	1 ASB1_HUMAN	Q9y576 mus musculus
3	481	20.5	336	1 ASB1_MOUSE	Q9wv74 mus musculus
4	296.5	12.7	643	1 MTR2_HUMAN	Q13614 homo sapien
5	288	12.3	391	1 MTR7_HUMAN	Q9y216 homo sapien
6	287	12.2	465	1 MTR6_HUMAN	Q9y217 homo sapien
7	283	12.1	669	1 MTR1_MOUSE	Q922c4 mus musculus
8	281	12.0	662	1 MTR1_HUMAN	Q13613 homo sapien
9	270.5	11.5	278	1 AS13_HUMAN	Q8wxk3 homo sapien
10	264	11.3	603	1 MTW1_HUMAN	Q13496 homo sapien
11	261.5	11.2	329	1 ASB5_HUMAN	Q8wxk0 homo sapien
12	258	11.0	603	1 MTW1_MOUSE	Q922c5 mus musculus
13	250	10.7	290	1 ASB9_MOUSE	Q912t8 mus musculus
14	246	10.5	323	1 AS11_HUMAN	Q8wxh4 homo sapien
15	243.5	10.4	518	1 ASB3_HUMAN	Q9y575 homo sapien
16	241	10.3	294	1 ASB3_HUMAN	Q96dx5 homo sapien
17	228.5	9.8	525	1 ASB3_MOUSE	Q9wv72 mus musculus
18	224.5	9.6	1862	1 ANK1_MOUSE	Q02357 mus musculus
19	224.5	9.6	1880	1 ANK1_HUMAN	P16157 homo sapien
20	215	9.2	1166	1 TNK2_HUMAN	Q9h2k2 homo sapien
21	210.5	9.0	688	1 YJ80 YEAST	P47147 saccharomyc
22	208.5	8.9	1327	1 TNK1_HUMAN	Q95271 homo sapien
23	205	8.7	4377	1 ANK3_HUMAN	Q12555 homo sapien
24	202.5	8.6	3924	1 ANK2_HUMAN	Q01484 homo sapien
25	202	8.6	768	1 YB23_HUMAN	Q91j77 homo sapien
26	199.5	8.5	288	1 ASB8_MOUSE	Q912t9 mus musculus
27	199.5	8.5	587	1 ASB2_HUMAN	Q96q27 homo sapien
28	194	8.3	583	1 AS15_MOUSE	Q8vhs6 mus musculus
29	193.5	8.3	429	1 AS10_HUMAN	Q8wxk3 homo sapien
30	190.5	8.1	1059	1 Y379_HUMAN	O15084 homo sapien
31	188.5	8.0	288	1 ASB8_HUMAN	Q9h765 homo sapien
32	187	8.0	410	1 V240 FOWPV	P14360 fowlpox vir
33	184.5	7.9	434	1 AS15_HUMAN	Q8wxk1 homo sapien

Query Match 68.2%; Score 1599; DB 1; Length 308;
Best Local Similarity 100.0%; Pred. No. 3e-118;
Matches 308; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	141	MNLMDITKIFSLQDPKKEEDTDTEKQALNAVYDNDSTYLDQLLRQRYKRFNSRSG	200
Db	1	MNLMDITKIFSLQDPKKEEDTDTEKQALNAVYDNDSTYLDQLLRQRYKRFNSRSG	60
Qy	201	MGVPGTPLRLAASGYHLSCLQVLLAHGADVDSLDVKAQTPFLTFVSHGHLCDCVRVLEAG	260
Db	61	MGVPGTPLRLAASGYHLSCLQVLLAHGADVDSLDVKAQTPFLTFVSHGHLCDCVRVLEAG	120

RESULT 1

AS12_HUMAN					
ID	AS12_HUMAN	STANDARD;	PRT;	308 AA.	
AC	Q8wxk4;				
DT	28-FEB-2003 (Rel. 41, Created)				
DT	28-FEB-2003 (Rel. 41, Last sequence update)				
DE	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Ankyrin repeat and SOCS box containing protein 12 (ASB-12).				
GN	ASB12.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RP	SEQUENCE FROM N.A.				
RA	Kile B.T., Nicola N.A.;				
RT	"SOCS box proteins.";				
RL	Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.				
CC	-!- SIMILARITY: Contains 5 ANK repeats.				
CC	-!- SIMILARITY: Contains 1 SOCS box domain.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; AF403030; AAL57349.1; -				
DR	Genew; HGNC:19763; ASB12.				
DR	InterPro; IPR002110; ANK.				
DR	Pfam; PF00023; ank; 5.				
DR	PRINTS; PR01415; ANKYRIN.				
DR	SMART; SM00248; ANK; 5.				
DR	PROSITE; PS50088; ANK_REPEAT; 4.				
DR	PROSITE; PS50297; ANK_REPEAT_REGION; 1.				
KW	ANK repeat; Repeat.				
FT	REPEAT 63 92 ANK 1.				
FT	REPEAT 96 125 ANK 2.				
FT	REPEAT 129 158 ANK 3.				
FT	REPEAT 171 200 ANK 4.				
FT	REPEAT 213 243 ANK 5.				
SQ	SEQUENCE 308 AA; 33829 MW; 80C47C544D03D8AC CRC64;				

Q915h8 fowlpox vir
Q915i6 fowlpox vir
Q914z4 fowlpox vir
Q92625 homo sapien
P59672 mus musculus
Q91zu0 mus musculus
Q9y574 homo sapien
Q9wv71 mus musculus
Q05921 mus musculus
Q96i34 homo sapien
Q08527 chromatium
Q96ns5 homo sapien

ALIGNMENTS

QY 261 ASPGSGIYNNCSVLTAAADGVAIIQELLDHGAENAVKAKLPVWASNTASCSGPLYLAA 320
 Db 121 ASPGSGIYNNCSVLTAAADGVAIIQELLDHGAENAVKAKLPVWASNTASCSGPLYLAA 180
 QY 321 VVGHLCDFRLLHLHGADPDYNTCDQGLLARVPRTLLLEICLHNCPEYIQLLIDFGAN 380
 Db 181 VVGHLCDFRLLHLHGADPDYNTCDQGLLARVPRTLLLEICLHNCPEYIQLLIDFGAN 240
 QY 381 IYLPSSLDLTSDDDKGIALLQARATPRSLLSQVRLVVRALCQAGQQAQINQLDIPM 440
 Db 241 IYLPSSLDLTSDDDKGIALLQARATPRSLLSQVRLVVRALCQAGQQAQINQLDIPM 300
 QY 441 LISYKQ 448
 Db 301 LISYKQ 308

RESULT 2

ASB1_HUMAN STANDARD; PRT; 335 AA.
 AC Q9Y576; Q9ULS4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Ankyrin repeat and SOCS box containing protein 1 (ASB-1).
 GN ASB1 OR KIAA1146.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20564172; PubMed=11111040;
 RA Kile B.T., Viney E.M., Willson T.A., Brodnicki T.C., Cancilla M.R.,
 RA Herlihy A.S., Croker B.A., Baca M., Nicola N.A., Hilton D.J.,
 RA Alexander W.S.;
 RT "Cloning and characterization of the genes encoding the ankyrin repeat
 RT and SOCS box-containing proteins Asb-1, Asb-2, Asb-3 and Asb-4.";
 RL Gene 258:31-41(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulláhy S.J.,
 RA Bosak S.A., McSwain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Murny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Nadeau A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE OF 65-335 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20039618; PubMed=10574461;
 RA Hirosewa M., Nagase T., Ishikawa K.-I., Kikuno R., Nomura N.,
 RA Ohara O.;
 RT "Characterization of cDNA clones selected by the GeneMark analysis
 RT from size-fractionated cDNA libraries from human brain.";
 RL DNA Res. 6:329-336(1999).

CC -!- SIMILARITY: Contains 6 ANK repeats.
 CC -!- SIMILARITY: Contains 1 SOCS box domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC
 CC EMBL; AF156777; AAD41894.1; -;
 DR EMBL; BC014528; AAH14528.1; -;
 DR EMBL; AB032972; EAA86460.1; -;
 DR MIM; HGNC:16011; ASB1.
 DR MIM; 605758; -;
 DR HSSP; Q00420; IAWC.
 DR GO; GO:0005622; C:intracellular; IC.
 DR GO; GO:0042036; P:negative regulation of cytokine biosynthesis; NAS.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR001496; SOCS.
 DR Pfam; PF00023; ank; 6.
 DR SMART; SM00248; ANK; 5.
 DR PROSITE; PS50088; ANK REPEAT; 3.
 DR PROSITE; PS50297; ANK REP REGION; 1.
 DR PROSITE; PS50225; SOCS; 1.
 KW ANK repeat; Repeat.
 FT REPEAT 36 68 ANK 1.
 FT REPEAT 77 106 ANK 2.
 FT REPEAT 110 139 ANK 3.
 FT REPEAT 143 172 ANK 4.
 FT REPEAT 191 220 ANK 5.
 FT REPEAT 235 265 ANK 6.
 FT DOMAIN 286 335 SOCS BOX.
 SQ SEQUENCE 335 AA; 37014 MW; 0843C96ADIAF60D9 CRC64;
 Query Match 21.3%; Score 499; DB 1; Length 335;
 Best Local Similarity 39.1%; Pred. No. 6.9e-32;
 Matches 119; Conservative 57; Mismatches 102; Indels 26; Gaps 9;
 QY 166 EKQALNQAVYDNDSTYTLQDLRQERYKRFNSRSGW---GVFGTPLRLAASYGHLSCLOV 222
 Db 37 EDTRLHDAAYVGLDTLRLSLQEEYSRINEKSVCCGWLPCTPLRIAATAAGHSCVDF 96
 QY 223 LLAHGADVDSLDVKAQTPLFTAVSHGHLDCTVRLLEAGASPGSGSYNNCSVLTAAADGA 282
 Db 97 LTRKGAEDLVDKGQTALYAVAVNGHLESTQILLEAGADPNGRHRSRTPVYHASRVGR 156
 QY 283 VAILQELLDHGAENAVKAKL-----PVWASNTAS---CSGPLYLAAYVGHLCDFRLLHL 334
 Db 157 ADILKALIRYGDVVDVNNHLLTDPVQPRFSRLTSLVVC--PLYISAAVHNLCQFRLLHL 214
 QY 335 GADPDYNC---TDQGLLARVPRTLLLEICLHNCPEYIQLLIDFGANIYL---PSLS 387
 Db 215 GANPDFNCNGPVNTQGYR--GSPGCVMDVLAHRGCEAAAFVSLVFEFGANLVKWSIG 272
 QY 388 LDLTSQ---DKGIALLQARATPRSLLSQVRLVVRALCQAGQQAQINQLDIPMLISY 444
 Db 273 PESRRGRKVDPEALQVFKEARSPVRTLCLCRVAVRRALGK-HRLHLPLSLPLPPIKFK 331
 QY 445 LKHQ 448
 Db 332 LUHE 335
 RESULT 3
 ASB1_MOUSE
 ID ASB1_MOUSE STANDARD; PRT; 336 AA.
 AC Q9WV74;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Ankyrin repeat and SOCS box containing protein 1 (ASB-1).


```
Query Match 12.7%; Score 296.5; DB 1; Length 643;
Best Local Similarity 25.4%; Pred. No. 1.3e-15; Indels 119; Gaps 17;
Matches 100; Conservative 57; Mismatches 118;

QY 3 ESKWKIDPISDFGRGIMRNWITTDANRNYEICSYPPIEVPKSVTLGTGVGSSKFR 62
   :|||||: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 203 ENGKLYDPLLEYRRGIPNESWRITKINERYELCDYPAJJVVVPANIPDEELKRVASR 262

QY 63 SKERVPLSVLYKENNAACRCSQPLSGPY-TRCVDDLELLRAISQTNPGSOFMYVVDPR 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 263 SRGRIPLSWIHPESQATITRCSQPMVGSGRKSDEKYLQAIMDSNAQSHKIFIFDAR 322

QY 122 PKI-----WHPLVLMRIVLQAKNLM--DITKIFSLLOPDKEE--- 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 323 PSNVAVANKAGGYESEDAYQNAELVF---LDIHNHVMSRLKKEIVYVNIETHW 379

QY 160 ----EDT-----DTEKQALNQAAYDNDVY-----TLDOLLROERYK 192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 380 LSNLESTHLEHKLILAGALRIADKVESGKTSVVVHCSGDGWDRTAQLTSLAMLMLDGY 439

QY 193 RFINS-----RSGWGVPGTPLRAASYHGLSCLQVLLAHG-----ADVSLDVKAQPLFT 243
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 440 RTIRGFEVLVEKEW-----LSFGHR--FQLRVGHGDKNHADAD-----RSPVEL 481

QY 244 AVSHGHLCVRLVLEAGASPGGSIVNCSFVLTAARDGAVAILQELLDH----- 292
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 482 Q-----FIDCVQWQTR--QFTAFEFNE-----YFLITILDHLVSLCLFGFTFLC 522

QY 293 -----GAENYKAKPLVWNA---SNIASCSGPLY 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 523 NSEQOQGENLPRVTYSLASYNLSQLEDFTNPLY 556

RESULT 5
MTR7 HUMAN STANDARD; PRT; 391 AA.
AC Q9Y2L6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myotubularin related protein 7 (EC 3.1.3.48) (Fragment).
GN MTR7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98409499; PubMed=9736772;
RA Laporte J., Blondeau F., Buj-Bello A., Tentler D., Kretz C., Dahl N.,
RA Mandel J.-L.;
RT "Characterization of the myotubularin dual specificity phosphatase
RT gene family from yeast to human.";
RL Hum. Mol. Genet. 7:1703-1712(1998).
CC -!- FUNCTION: NOT KNOWN, COULD BE A TYROSINE-PHOSPHATASE.
CC MTR7.
CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphatase.
CC -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN BRAIN.
CC -!- FUNCTION: NOT KNOWN, COULD BE A TYROSINE-PHOSPHATASE.
CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein
CC tyrosine + phosphate.
CC -!- SIMILARITY: BELONGS TO THE MYOTUBULARIN-LIKE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF073482; AAC77820.1; -.
CC Genew; HGNC:7454; MTR7.
CC MIM; 603562; -.
CC -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN BRAIN.
CC -!- SIMILARITY: BELONGS TO THE MYOTUBULARIN-LIKE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF073482; AAC77820.1; -.
CC Genew; HGNC:7454; MTR7.
CC MIM; 603562; -.
CC GO; GO:0006470; P:protein amino acid dephosphorylation; TAS.

DR InterPro; IPR00387; TYR phosphatase.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
KW Hydrolase.
FT ACT_SITE 174 174 BY SIMILARITY.
FT NON_TER 1 1
SQ SEQUENCE 391 AA; 45437 MW; 61C6C104AEBE1C50 CRC64;

Query Match 12.3%; Score 288; DB 1; Length 391;
Best Local Similarity 65.4%; Pred. No. 3.1e-15;
Matches 53; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 44 IVPKSVTLGTGVGSSKFRKERVPLSVLYLYKENNAACRCSQPLSGFYTRCVDDLELL 103
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 LVVPSATAHIIVGSKFRSRRRFFVLSYYKDNHASICRSSQPLSGFSARCLEDEQMLQ 60

QY 104 AISQTNPGSQFMYVVDTRPKI 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 AIRKANPGSDFVYVVDTRPKL 81

RESULT 6
MTR6 HUMAN STANDARD; PRT; 465 AA.
AC Q9Y2L7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myotubularin related protein 6 (EC 3.1.3.48) (Fragment).
GN MTR6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98409499; PubMed=9736772;
RA Laporte J., Blondeau F., Buj-Bello A., Tentler D., Kretz C., Dahl N.,
RA Mandel J.-L.;
RT "Characterization of the myotubularin dual specificity phosphatase
RT gene family from yeast to human.";
RL Hum. Mol. Genet. 7:1703-1712(1998).
CC -!- FUNCTION: NOT KNOWN, COULD BE A TYROSINE-PHOSPHATASE.
CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein
CC tyrosine + phosphate.
CC -!- SIMILARITY: BELONGS TO THE MYOTUBULARIN-LIKE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF072928; AAC78841.1; -.
CC Genew; HGNC:7453; MTR6.
CC MIM; 603561; -.
CC GO; GO:0004722; F:protein serine/threonine phosphatase activity; NAS.
CC GO; GO:0006470; P:protein amino acid dephosphorylation; NAS.
DR InterPro; IPR00387; TYR phosphatase.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; FALSE_NEG.
KW Hydrolase.
FT ACT_SITE 180 180 BY SIMILARITY.
FT NON_TER 1 1
SQ SEQUENCE 465 AA; 53608 MW; D24518975AE75A2B CRC64;

Query Match 12.2%; Score 287; DB 1; Length 465;
Best Local Similarity 61.6%; Pred. No. 4.6e-15;
Matches 53; Conservative 16; Mismatches 17; Indels 0; Gaps 0;

QY 39 TYPPEIVVPKSVTLGTGVGSSKFRKERVPLSVLYLYKENNAACRCSQPLSGFYTRCVDD 98
```


Db 2 TYPELVVPIASKPIITVSSKFRGPFVLSVYHODKEAAICRCSQPLSGFSARCLD 61
Qy 99 ELLEAISQTPGSOQFMVVDTRPKI 124
Db 62 EHLQAISKANPVNRYMYVMDTRPKL 87

RESULT 7

MTRL_MOUSE
ID MTRL_MOUSE STANDARD; PRT; 669 AA.
AC Q922C4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myotubularin-related protein 1.
GN MTMRI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98409499; PubMed=9736772;
RA Laporte J., Blondeau F., Buj-Bello A., Tentler D., Kretz C., Dahl N.,
Mandel J.-L.;
RT "Characterization of the myotubularin dual specificity phosphatase
gene family from yeast to human";
RL Hum. Mol. Genet. 7:1703-1712(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Wiehe T., Zhao W., Herman G.E., Rosenthal A., Platzter M.;
RT "Comparative sequence analysis of the mouse Mtm locus and the
corresponding region of human Xq28";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: NOT KNOWN, COULD BE A TYROSINE-PHOSPHATASE.
CC -1- SIMILARITY: BELONGS TO THE MYOTUBULARIN-LIKE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
EMBL; AF073997; AAC7822.1; -
DR EMBL; AF125314; AAF22122.1; -
DR MGD; MGI:1858271; Mtmrl.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005886; C:plasma membrane; IDA.
DR GO; GO:0004438; F:phosphatidylinositol-3-phosphatase activity; IDA.
DR InterPro; IPR004182; GRAM dom.
DR InterPro; IPR000387; TYR_phosphatase.
DR Pfam; PF02893; GRAM; 1.
DR SMART; SM00568; GRAM; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PSS0056; TYR_PHOSPHATASE_2; 1.
KW Hydrolase.
FT ACT SITE 442 442 BY SIMILARITY.
SQ SEQUENCE 669 AA; 75313 MW; 1856792245F2D800 CRC64;

Query Match 12.18; Score 283; DB 1; Length 669;
Best Local Similarity 27.5%; Pred. No. 1.6e-14;
Matches 86; Conservative 42; Mismatches 93; Indels 92; Gaps 12;
Qy 4 SGWKLIDPIGDFGSMGIPNRTWITDANRNVICSTYPPPIVVPKSVTLGTGVGSSKFRS 63
Db 229 NGWKVDPVSEYKQGFNWSKLSKINSVEFCFTYPAIVVPTSVKDDDLKVAAPRA 288
Qy 64 KERPVLISLYKNNAAICRCSQPLSG - FYTRCVDDLELLEAISQTPGSOQFMVVDTR - 121
Db 289 KGRVPVLISWHIPESQAITTCRCSQPLGVGPNDRCKEDEKYLQTIMDANAQSHKLTFIDRQ 348

Qy 122 -----PKIWHFLVLMIRVLQAKNLM--DITKIFSLQPKDEE 159
Db 349 NSVADTNKAKGGYENESAYPN-----AEIJFLEIHNHVMRESLRKLEIVYPSIDE 401
Qy 160 ED-----TDTTEKQALNQAQVYDNDSY-----TLDQLLRQRE 189
Db 402 SHWLSNVDGTHWLEIVIRVLLAGAVRIADKIESGKTSVVIHCSGDWDRTSQLTSLAMLMD 461
Qy 190 RYKRFINS-----RSGWGVPGTFLRLAASGYHLSCLQVLLAHG-----ADVDSLDKVAQTP 240
Db 462 SYRTIKNGPEALIEKEW-----ISFGHFRALRV--GHGDDNHADAD-----RSP 503
Qy 241 LFTAVSHGHLDVCV 253
Db 504 IFLQ----FIDCV 512
RESULT 8
MTRL_HUMAN
ID MTRL_HUMAN STANDARD; PRT; 662 AA.
AC Q136I3; Q9UBX6; Q9UEM0; Q9UOD5;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myotubularin-related protein 1 (Fragment).
GN MTMRI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RX MEDLINE=99047532; PubMed=9828128;
RA Kioschis P., Wiemann S., Heiss N.S., Francis F., Goetz C., Poustka A.,
Taudien S., Platzter M., Wiehe T., Beckmann G., Weber J., Nordsiek G.,
Rosenthal A.;
RT "Genomic organization of a 225-kb region in Xq28 containing the gene
for X-linked myotubular myopathy (MTM1) and a related gene (MTMRI).";
RL Genomics 54:256-266(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Brain;
RA Kioschis P., Wiemann S., Francis F., Goetz C., Poustka A., Taudien S.,
Platzter M., Wiehe T., Beckmann G., Weber J., Nordsiek G.,
Rosenthal A.;
RT "Ancient genomic duplication within the myotubular myopathy locus
(MTM1) in human Xq28";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 111-607 FROM N.A.
RX MEDLINE=98409499; PubMed=9736772;
RA Laporte J., Blondeau F., Buj-Bello A., Tentler D., Kretz C., Dahl N.,
Mandel J.-L.;
RT "Characterization of the myotubularin dual specificity phosphatase
gene family from yeast to human";
RL Hum. Mol. Genet. 7:1703-1712(1998).
RN [4]
RP SEQUENCE OF 401-532 FROM N.A.
RX MEDLINE=96225444; PubMed=8640223;
RA Laporte J., Hu L.J., Kretz C., Mandel J.-L., Kioschis P., Coy J.,
Klauck S.M., Poustka A., Dahl N.;
RT "A gene mutated in X-linked myotubular myopathy defines a new
putative tyrosine phosphatase family conserved in yeast.";
RL Nat. Genet. 13:175-182(1996).
CC -1- FUNCTION: NOT KNOWN, COULD BE A TYROSINE-PHOSPHATASE.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q136I3-1; Sequence=Displayed;
CC Name=1A;
CC IsoId=Q136I3-2; Sequence=VSP_005169, VSP_005170;
CC -1- SIMILARITY: BELONGS TO THE MYOTUBULARIN-LIKE FAMILY.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF002223; ; NOT ANNOTATED_CDS.
CC EMBL; AJ224979; CAAL2271.1; -.
CC EMBL; AF057354; AAD40368.1; -.
CC EMBL; U58032; AAC79117.1; -.
CC Genew; HGNC:7449; MTMR1.
CC MTM; 300171; -.
CC GO; GO:0004722; F:protein serine/threonine phosphatase activity; ISS.
CC GO; GO:0004725; F:protein tyrosine phosphatase activity; ISS.
CC GO; GO:0006470; P:protein amino acid dephosphorylation; ISS.
CC InterPro; IPR004182; GRAM dom.
CC InterPro; IPR000387; TYR_phosphatase.
CC Pfam; PF02893; GRAM; 1.
CC SMART; SM00568; GRAM; 1.
CC PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
CC PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.
CC Hydrolase; Alternative splicing.
KW NON TER 1 1
FT ACT_SITE 435 435 BY SIMILARITY.
FT DOMAIN 2 7 POLY-ALA.
FT DOMAIN 11 14 POLY-GLY.
FT DOMAIN 642 645 POLY-SER.
FT VARSPPLIC 550 565 DVTYKTSLSMYSINQ -> AMGAGTQARGSLRSR (in isoform 1A).
FT VARSPPLIC 566 662 Missing (in isoform 1A).
FT CONFLICT 538 538 /FTID=VSP_005170.
FT SEQUENCE 662 AA; 74275 MW; 4FB8FCOA5878FC2C CRC64;
Query Match 12.0%; Score 281; DB 1; Length 662;
Best Local Similarity 27.5%; Pred.No. 2.2e-14;
Matches 86; Conservative 42; Mismatches 93; Indels 92; Gaps 12;
QY 4 SGKKLIDPISDFGRMGIPNRRWTTITDARNYEICSTYPPPIVVKSVTLGTGVGSSKFRS 63
DB 222 NGKVVDPVSEYKQGLPNESKINSNYEPCDTYPAIIIVPTSVKDDLSKVAAPRA 281
QY 64 KERVPVLSLYKENNAIICRCSPLSG-FYTRCVDDDELLEALISQNPQSGQFVVVDTR- 121
DB 282 KGRVPVLSWIHPESQATITRCSQPLVGPNDKCKEKEYLQTIMDANAQSHKLIIFDARQ 341
QY 122 -----PKIWHFLVLMIRIVLQAKNLM--DITKIFSLQPDKEE 159
DB 342 NSVADTNKTKGGYSESAYPN-----AELVFLEIHNIHVNRSLRKLKEIVPSIDE 394
QY 160 -----EDTTEEQALNAQVYDNDY-----TLDQLLRQE 189
DB 395 ARWLSNVDGTHMLEYIRMLLAGAVRADTKIESKTSVWVHCSDGWRDTAQLTSLAMLMD 454
QY 190 RYKRFINS-----RSQGVGPGTPLRLAASVGHLSCLQVLLAHG-----ADVSLDVKQAQTP 240
DB 455 SYRTIKGPFETLVEKEW-----ISFGHRFALRV--GHGNDNHADAD-----RSP 496
QY 241 LFTAVSHGHLDVCV 253
DB 497 IFLQ-----FVDCV 505
RESULT 9
ID AS13 HUMAN STANDARD; PRT; 278 AA.
AC Q8WXK3; Q9SEP7; Q9H8Z1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)

```

```

DT DE Ankyrin repeat and SOCS box containing protein 13 (ASB-13).
GN ASB13.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Kile B.T., Nicola N.A.;
RT "SOCS box proteins.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Isegai T., Ota T., Hayaehi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuna M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDD human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
TI TISSUE=Uterus;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Millay S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- ALTERNATIVE PRODUCTS;
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q8WXK3-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8WXK3-2; Sequence=VSP_000272, VSP_000273;
CC Note=No experimental confirmation available;
CC -!- SIMILARITY: Contains 6 ANK repeats.
CC -!- SIMILARITY: Contains 1 SOCS box domain.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 14.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF403031; AAL57350.1; ALT_FRAME.
CC EMBL; AK023195; BAB14456.1; -.
CC EMBL; BC012056; AAH12056.1; -.
CC Genew; HGNC:19765; ASB13.
CC InterPro; IPR002110; ANK.
CC InterPro; IPR001496; SOCS.

```


DR EMBL; U46024; AAC12865.1; -.
 DR EMBL; AF020676; AAC12865.1; JOINED.
 DR EMBL; AF020664; AAC12865.1; JOINED.
 DR EMBL; AF020665; AAC12865.1; JOINED.
 DR EMBL; AF020666; AAC12865.1; JOINED.
 DR EMBL; AF020667; AAC12865.1; JOINED.
 DR EMBL; AF020668; AAC12865.1; JOINED.
 DR EMBL; AF020669; AAC12865.1; JOINED.
 DR EMBL; AF020670; AAC12865.1; JOINED.
 DR EMBL; AF020671; AAC12865.1; JOINED.
 DR EMBL; AF020672; AAC12865.1; JOINED.
 DR EMBL; AF020673; AAC12865.1; JOINED.
 DR EMBL; AF020674; AAC12865.1; JOINED.
 DR EMBL; AF020675; AAC12865.1; JOINED.
 DR EMBL; BC030779; AAH30779.1; -.
 DR Genew; HGNC:7448; MTWL.
 DR MIM; 300415; -.
 DR MIM; 310400; -.
 DR GO; GO:0004722; P:protein serine/threonine phosphatase activity; NAS.
 DR GO; GO:0008151; P:cell growth and/or maintenance; TAS.
 DR GO; GO:0007517; P:muscle development; TAS.
 DR GO; GO:0006470; P:protein amino acid dephosphorylation; NAS.
 DR InterPro; IPR0041182; GRAM Dom.
 DR InterPro; IPR000387; TYR_phosphatase.
 DR Pfam; PF02893; GRAM; 1.
 DR SMART; SM00568; GRAM; 1.
 DR PROSITE; PS00383; TYR PHOSPHATASE_1; 1.
 DR PROSITE; PS00566; TYR PHOSPHATASE_2; 1.
 KW Disease mutation; Hydrolyase.
 FT ACT_SITE 375 375
 FT VARIANT 47 47
 FT VARIANT 69 69
 FT VARIANT 70 70
 FT VARIANT 87 87
 FT VARIANT 179 179
 FT VARIANT 184 184
 FT VARIANT 189 189
 FT VARIANT 198 198
 FT VARIANT 205 205
 FT VARIANT 225 225
 FT VARIANT 229 229
 FT VARIANT 241 241
 FT VARIANT 241 241
 FT VARIANT 264 264
 FT VARIANT 294 294
 FT VARIANT 317 317
 FT VARIANT 376 376
 FT VARIANT 378 378
 FT VARIANT 397 397
 FT VARIANT 402 402
 FT VARIANT 404 404
 FT VARIANT 406 406

FT VARIANT 420 420 /FTID=VAR_006403.
 FT S -> SFIQ (IN XLMTM; SEVERE).
 FT /FTID=VAR_009221.
 FT R -> Q (IN XLMTM; SEVERE).
 FT /FTID=VAR_006404.
 FT R -> RFIQ (IN XLMTM; SEVERE).
 FT /FTID=VAR_006405.
 FT D -> N (IN XLMTM).
 FT /FTID=VAR_006406.
 FT D -> N (IN XLMTM).
 FT /FTID=VAR_006407.
 FT H -> P (IN XLMTM).
 FT /FTID=VAR_006408.
 FT W -> R (IN XLMTM; MILD).
 FT /FTID=VAR_006409.
 FT K -> N (IN XLMTM; SEVERE).
 FT /FTID=VAR_009222.
 FT E -> K (IN REF. 3).
 FT CONFLICT 410 410
 FT SEQUENCE 603 AA; 69931 MW; BE970F2471957C0 CRC64;
 Query Match 11.3%; Score 264; DB 1; Length 603;
 Best Local Similarity 26.7%; Pred. No. 4.2e-13;
 Matches 80; Conservative 46; Mismatches 106; Indels 68; Gaps 12;
 QY 5 GWKLIDPISDFGRMGIPNRNWTITDANRNEYICSTYPPPEIVVPKSVTLGTVVGSSEKFRSK 64
 DB 163 GWTVPVPEYRQGLPNHWRITINKCYELCDYTPALLVVPYRASDDDDLRVATFRSR 222
 QY 65 ERVPVLSYLYKENNAICRCQSPLSGFY-TRCVDELLEALISQTPGSGQFMYVVDTRPK 123
 DB 223 NRIPVLSWIHPENKTVIVRCQSPLVGMGSKRNDKDYLDVIRETNKQISKLTIVDAREPS 282
 QY 124 I-----WHFLVLMIRIVLOLAKNLM--DITKIFSLLOPKKE----- 159
 DB 283 VNAVANKATGGYESSDDAYHNAELFF---LDIHNTHVRESLKVKVDIVPNVESHWLS 339
 QY 160 --EDTTEE--KQALNQAQVYNDSTYLDQLLRQERYKRFINSRSGWGVGTPLRLA---- 211
 DB 340 SLESTHWLEHIKLVLTGAIQVADK-----VSSGSSSVLVHSCDGDWDRTAQLTSLAAML 393
 QY 212 -----ASYGHLSCLOVLAHGADVDLSLVKQAQTPLFTAVSHGHLDVCV 253
 DB 394 DSFYRSIEGFEILVQKEWISFGHKFASRI--GHG-DKNHTDAD-RSPIPLQ----FIDCV 445
 RESULT 11
 ASB5 HUMAN
 ID ASB5 HUMAN STANDARD; PRT; 329 AA.
 AC Q8WVK0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Ankyrin repeat and SOCS box containing protein 5 (ASB-5).
 GN ASB5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kile B.T., Hilton D.J., Nicola N.A.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: Contains 6 ANK repeats.
 CC -!- SIMILARITY: Contains 1 SOCS box domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use: by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischnann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli P., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wuyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- SIMILARITY: Contains 6 ANK repeats.
CC -!- SIMILARITY: Contains 1 SOCS box domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF398970; AAK97492.1; -;
DR EMBL; AK008978; BAB24293.1; -;
DR MGD; MG1:1916549; ASB9.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001496; SOCS.
DR Pfam; PF00023; ank; 6.
DR SMART; SM00248; ANK; 6.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT; 1.
DR PROSITE; PS50225; SOCS; 1.
KW ANK repeat; Repeat.
FT REPEAT 31 60 ANK 1.
FT REPEAT 64 93 ANK 2.
FT REPEAT 97 126 ANK 3.
FT REPEAT 129 158 ANK 4.
FT REPEAT 162 191 ANK 5.
FT REPEAT 194 223 ANK 6.
FT DOMAIN 236 290 SOCS BOX.
FT CONFLICT 275 275 C -> S (IN REF. 2).
SQ SEQUENCE 290 AA; 31672 MW; 25F9B9A086951B856 CRC64;

Query Match 10.7%; Score 250; DB 1; Length 290;
Best Local Similarity 29.7%; Pred. No. 2e-12;
Matches 88; Conservative 36; Mismatches 112; Indels 60; Gaps 10;

QY 170 LNQAVYDNDSTYLDQLRQERYKRFINSRGV-----PGTPLRLAASYGHLSCQLVYL 223
DB 36 LQDAATHGCLLTNRNLSQ-----GMPVNIITADHVSPLPHEACLRGHLSCASVYL 84
QY 224 LAHGADVDSDLVKATQPLFTAVSHGLDCRVLLLEAGSPGSIYNNCSFVLTAARDGAV 283
DB 85 LSHGAQVQWMTDWTPLFNACVSGQDCVNLQHGATPHPET-ELASPIHEAAKRGV 143
QY 284 AILQELLDHGAANVAKLPLVWASNIASCSGPLYLAAYVGHLDPCFRLLLHGADEPDYNT 343
DB 144 KCIESLAAGANIDY-----NISHLGTPLYVACKNQVACAKKLLSGVS-----V 189
QY 344 DQG-----LLARVPRRTLLEICLHNCPEYIQLLDGFIYPLSL-----LDLT 391
DB 190 NQGGKGLDSPHVARM-----SSVELVHLLMDFGANAQAKNAGKRPVDIV 235

QY 392 SODDKGIALLLQARATPRSLLSQVRLVVRALCQAGQPOAQNQLDIPPLMISYLKH 447
DB 236 PLESPLIQIFLQNEG-POSRLQCLRLINKCF-GIRQHHKICELLPLDKRFLH 289
RESULT 14
ID AS11 HUMAN STANDARD; PRT; 323 AA.
AC Q8WYH4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ankyrin repeat and SOCS box containing protein 11 (ASB-11).
GN ASB11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kile B.T., Hilton D.J., Nicola N.A.;
RT "SOCS box proteins.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 6 ANK repeats.
CC -!- SIMILARITY: Contains 1 SOCS box domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF425642; AAL60519.1; -;
DR Genew; HGNC:17186; ASB11.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001496; SOCS.
DR Pfam; PF00023; ank; 6.
DR SMART; SM00248; ANK; 6.
DR PROSITE; PS50088; ANK_REPEAT; 6.
DR PROSITE; PS50297; ANK_REPEAT; 1.
DR PROSITE; PS50225; SOCS; 1.
KW ANK repeat; Repeat.
FT REPEAT 64 93 ANK 1.
FT REPEAT 97 126 ANK 2.
FT REPEAT 130 159 ANK 3.
FT REPEAT 162 191 ANK 4.
FT REPEAT 195 224 ANK 5.
FT REPEAT 227 256 ANK 6.
FT DOMAIN 273 323 SOCS BOX.
SQ SEQUENCE 323 AA; 35367 MW; F95A40184C5BBF4E CRC64;

Query Match 10.5%; Score 246; DB 1; Length 323;
Best Local Similarity 29.9%; Pred. No. 4.7e-12;
Matches 82; Conservative 44; Mismatches 100; Indels 48; Gaps 9;

QY 206 TPLRLAASYGHLSCQLVLLAHGADVDSDLVKATQPLFTAVSHGLDCRVLLLEAGSPG 265
DB 67 SPLHEAAQGRLLALKTLIAQGVNVLTVIRVSSLHEACLRGHLSCASVYL 126
QY 266 SIYNNCSFVLTAARDGAVAILQELLDHGAANVAKLPLVWASNIASCSGPLYLAAYVGH 325
DB 127 VTVHGATPLFNACSGSACVNVLLFEGAKAQLEVLH-----ASPIHEAVKRGH 176
QY 326 DCFRLLLHGADEPDYNT-----CTDQGLLARVPRRTLLEI-----CL 362
DB 177 ECWEILLANNVNIHVEVQLGTPLFVACTYQ-----RVCVKKLLGASVDHGQWLDTEL 232
QY 363 H---HNCPEYIQLLDGFIYPLSL-----SLDITS-ODDKGIALLLQARATPRSLLSQ 414
DB 233 HAAARQSNVEVTHLLTDYGANLKRENAQKGSALDIAAPKSSVEQALL--REGPPALSQ 290

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 1, 2003, 07:44:47 ; Search time 92 Seconds
(without alignments)
2154.142 Million cell updates/sec

Title: US-09-941-831a-20
Perfect score: 2343
Sequence: 1 MRESGWKLDPISDFGRMGI.....QAINQLDIPPLISLYLKHQL 449

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DRV=xlpl
-Q=/cgn2_1/USPTO.spool_p/US09941831/runat_01122003_073528_7855/app_query.fasta_1.647
-DB=issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09941831 -CGN1_1_85=runat_01122003_073528_7855 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	480	20.5	2019	4	US-09-302-769-24
2	264	11.3	3431	4	US-09-155-078-1
3	261.5	11.2	1206	4	US-09-620-312D-34
4	244.5	10.4	401	3	US-09-221-298-28
5	233.5	10.0	419	4	US-09-302-769-26
6	219	9.3	3498	4	US-09-350-982C-4
7	219	9.3	4512	4	US-09-350-982C-3
8	213.5	9.1	4431	3	US-09-196-387-7
9	213.5	9.1	4491	4	US-09-841-835-7
10	209	8.9	4657	3	US-09-196-387-9
11	209	8.9	4657	4	US-09-841-835-9
12	208.5	8.9	4134	3	US-09-196-387-1

13	208.5	8.9	4134	4	US-09-841-835-1	Sequence 1, Appli
14	207	8.8	1712	4	US-09-620-312D-1101	Sequence 1101, Ap
15	205	8.7	14770	4	US-09-220-132-30	Sequence 30, Appl
16	202	8.6	3454	3	US-09-082-089-1	Sequence 1, Appli
17	201.5	8.6	909	1	US-09-031-485-25	Sequence 25, Appl
18	201.5	8.6	909	1	US-09-031-485-26	Sequence 26, Appl
19	201.5	8.6	909	1	US-08-847-429A-25	Sequence 25, Appl
20	201.5	8.6	909	1	US-08-847-429A-26	Sequence 26, Appl
21	201.5	8.6	909	3	US-09-065-474-25	Sequence 25, Appl
22	201.5	8.6	909	3	US-09-065-474-26	Sequence 26, Appl
23	201.5	8.6	909	4	US-09-557-034-25	Sequence 25, Appl
24	201.5	8.6	909	4	US-09-557-034-26	Sequence 26, Appl
25	201.5	8.6	911	1	US-09-031-485-22	Sequence 22, Appl
26	201.5	8.6	911	1	US-09-031-485-24	Sequence 24, Appl
27	201.5	8.6	911	1	US-08-847-429A-22	Sequence 22, Appl
28	201.5	8.6	911	1	US-08-847-429A-24	Sequence 24, Appl
29	201.5	8.6	911	3	US-09-065-474-22	Sequence 22, Appl
30	201.5	8.6	911	3	US-09-065-474-24	Sequence 24, Appl
31	201.5	8.6	911	4	US-09-557-034-22	Sequence 22, Appl
32	201.5	8.6	911	4	US-09-557-034-24	Sequence 24, Appl
33	201.5	8.6	1056	3	US-09-065-474-138	Sequence 138, App
34	201.5	8.6	1056	3	US-09-065-474-140	Sequence 140, App
35	201.5	8.6	1056	4	US-09-557-034-138	Sequence 138, App
36	201.5	8.6	1056	4	US-09-557-034-140	Sequence 140, App
37	201.5	8.6	5235	1	US-09-031-485-35	Sequence 35, Appl
38	201.5	8.6	5235	1	US-09-031-485-36	Sequence 36, Appl
39	201.5	8.6	5235	1	US-08-847-429A-35	Sequence 35, Appl
40	201.5	8.6	5235	1	US-08-847-429A-36	Sequence 36, Appl
41	201.5	8.6	5235	3	US-09-065-474-35	Sequence 35, Appl
42	201.5	8.6	5235	3	US-09-065-474-36	Sequence 36, Appl
43	201.5	8.6	5235	4	US-09-557-034-35	Sequence 35, Appl
44	201.5	8.6	5235	4	US-09-557-034-36	Sequence 36, Appl
45	201.5	8.6	5503	1	US-09-031-485-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1
US-09-302-769-24
; Sequence 24, Application US/09302769
; Patent No. 6323317
; GENERAL INFORMATION:
; APPLICANT: HILTON, Douglas J
; APPLICANT: ALEXANDER, Warren S
; APPLICANT: VINEY, Elizabeth M
; APPLICANT: WILLSON, Tracey A
; APPLICANT: RICHARDSON, Rachael T
; APPLICANT: STARR, Robyn
; APPLICANT: NICHOLSON, Sandra E
; APPLICANT: METCALF, Donald
; APPLICANT: NICOLA, Nicolas A
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
; FILE REFERENCE: 109762
; CURRENT APPLICATION NUMBER: US/09/302,769
; CURRENT FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 08/962,560
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 24
; LENGTH: 2019
; TYPE: DNA
; ORGANISM: Mouse
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (2000)
; OTHER INFORMATION: n is unsure
US-09-302-769-24

Alignment Scores:
Pred. No.: 3,98e-48 Length: 2019
Score: 480.00 Matches: 114
Percent Similarity: 57.7% Conservative: 58

Best Local Similarity: 38.26% Mismatches: 104
Query Match: 20.49% Indels: 22
DB: 4 Gaps: 8

US-09-941-831A-20 (1-449) x US-09-302-769-24 (1-2019)

```
QY 170 LeuAsnGlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeuLeuArgGlnGlu 189
Db 167 CTCATGATGACGCTATGTAGGGACCTCCAGACCTCAGAACCTACTGCAAGAGGAG 226
QY 190 ArgTyrIysArgPheIleAsnSerArgSerGlyTrp-----GlyValProGlyThr 206
Db 227 AGCTACCGGACCGCATCAATGAGAGTCTGTCTGGTGTGCGCTGGCTTCCCTGCACA 286
QY 207 ProIeuArgLeuAlaIleAsnSerTyrGlyHisIeuSerCysLeuGlnValLeuAlaHis 226
Db 287 CCCTAGGATGACGACCACTGAGGACCATGGGAACCTGTGTGACCTTCCTCATACGCAAA 346
QY 227 GlyAlaAspValAspSerLeuAspValIysAlaGlnThrProLeuPheThrAlaValSer 246
Db 347 GGGGCCAGGTGACCTGGTGGATGTCAGGGCAGACTCCCTGTATGGCTGTAGTG 406
QY 247 HisGlyHisIeuAspCysValArgValLeuLeuGluAlaGlyAlaSerProGlyGlySer 266
Db 407 AACGGGCACCTTGGAGACCACTGAGATCTTTTGGAGCTGGTCTGTATCCCAACGGCAGC 466
QY 267 IleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaValAlaIleLeu 286
Db 467 CGGCACCAACGACCACTCTCTGTACCATGCTTGTGTGGTGTAGGACGACATCCTG 526
QY 287 GlnGluLeuLeuAspHisGlyAlaGluAlaAsnValIysAlaIysLeu----- 302
Db 527 AAGGCTCTTATCAGTATGGGCGAGATGTGATGTCAACCATCATCTGAATTCGACACC 586
QY 303 ---ProValTrpAlaSerAsnIleAlaSer-----CysSerGlyProLeuTyrLeu 318
Db 587 CGGCCCTTTTTCACGGCGGCTAACCTCTCTGGTGTCTGT-----CCTCTATACATC 640
QY 319 AlaAlaValTyrGlyHisIeuAspCysPheArgLeuLeuLeuLeuHisGlyAlaAspPro 338
Db 641 AGTGTGCTCTACCATCACTTCACTGCTTCAAGTGTCTTCAAGCTGCTTCAAGCTGCT 700
QY 339 AspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgValProArg-----ProArgThr 356
Db 701 GACTTCATTCGAATGCCCTGTCAACCCAGGAGTTCACAGGGGATCCCTCGGGTGT 760
QY 357 LeuLeuGluIleCysLeuHisHisAsnCysGluProGluTyrIleGlnLeuLeuIleAsp 376
Db 761 GTCATGATGCTCTCTCGCCCATGGCTGTGAAGCAGCCTTCGTGAGTCTGTGGTAGAG 820
QY 377 PheGlyAlaAsnIleTyrLeu-----ProSerLeuSerLeuAspLeuThrSerGln 393
Db 821 TTGGAGCCCAACCTGAACCTGTGTGAAGTGGGAATCCCTGGGCCACAGCAAGAGCAGA 880
QY 394 -----AspAspIysGlyIleAlaLeuLeuLeuGlnAlaArgAlaThrProArgSer 410
Db 881 AGAAGATGATCCTGAGCCCTTGCAGGCTTTAAAGAGCCACAGATATTCACGAGCC 940
QY 411 LeuLeuSerGlnValArgLeuValValArgAlaLeuCysGlnAlaGlyGlnProGln 430
Db 941 TTCTGAGTTTGTGCGGGTGGCTGTGAGAAGAGCTCTTGGCAAAATAC---CGACTGCAT 997
QY 431 AlaIleAsnGlnLeuAspIleProProMetLeuIleSerTyrLeuIysHisGln 448
Db 998 CTGGTTCCCTCGCTGCGCTGCCAGACCCCAATAAAGAGTTTGTCTTATGAG 1051
```

RESULT 2

US-09-155-078-1
; Sequence 1, Application US/09155078
; Patent No. 6312688
; GENERAL INFORMATION:
; APPLICANT: POUSTKA, Annemarie
; APPLICANT: KIOSCHIS, Petra

; APPLICANT: LAPORTE, Jocelyn
; APPLICANT: HU, Ling Jia
; APPLICANT: MANDEL, Jean Louis
; APPLICANT: DAHL, Niklas
; TITLE OF INVENTION: TYROSINE-PHOSPHATASE-RELATED PROTEIN
; FILE REFERENCE: 012627-010
; CURRENT APPLICATION NUMBER: US/09/155,078
; CURRENT FILING DATE: 1999-03-02
; EARLIER FILING DATE: 1997-03-21
; EARLIER APPLICATION NUMBER: PCT/DE97/00592
; EARLIER FILING DATE: 1996-03-21
; EARLIER APPLICATION NUMBER: DE 196 11 234.6
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatenIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3431
; TYPE: DNA
; ORGANISM: German-type microorganism & cell cul.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1863)
US-09-155-078-1

Alignment Scores:
Pred. No.: 2,4e-21 Length: 3431
Score: 264.00 Matches: 80
Percent Similarity: 42.00% Conservative: 46
Best Local Similarity: 26.67% Mismatches: 106
Query Match: 11.27% Indels: 68
DB: 4 Gaps: 12

US-09-941-831A-20 (1-449) x US-09-155-078-1 (1-3431)

```
QY 5 GlyTrpIysLeuIleAspProIleSerPheGlyArgMetGlyIleProAsnArgAsn 24
Db 541 GGATGGACAGTTTACATCCAGTGGGAAGAAATACAGGAGCGAGGCTTGCCCAATCACCAT 600
QY 25 TrpThrIleThrAspAlaAsnArgAsnTyrGluIleCysSerThrTyrProGluIle 44
Db 601 TCGAGAATAAATCTTTTAAATAAGTGTCTATGAGCTCTGTGACACTTACCTCTCTTTTG 660
QY 45 ValValProIysSerValThrLeuGlyThrValValGlySerSerIysPheArgSerIys 64
Db 661 GTGGTTTCGTATCGTCTGCTCAGATGATGATGATGATGATGATGATGATGATGATGAT 720
QY 65 GluArgValProValLeuSerTyrIleTyrIysGluAsnAsnAlaIleCysArgCys 84
Db 721 AATCGAATTCAGTCTGTCTGATGGATTCTCCAGAAAATAAGACGGTCTATTGTGGCTGC 780
QY 85 SerGlnProLeuSerGlyPheTyr---ThrArgCysValAspAspGluLeuLeuGlu 103
Db 781 AGTCAGCTCTTGTCTGGTATGATGCGGAACGAAATAAAGATGATGAGAAATATCTCGAT 840
QY 104 AlaIleSerGlnThrAsnProGlySerGlnPheMetTyrValValAspThrArgProIys 123
Db 841 GTTATCAGGAGACTTAATAACAAATTTCTAAACTCACCATTATGATGCAAGACCCAGC 900
QY 124 Ile-----TrpHis 126
Db 901 GTAAATGAGTGGCCCAACAGGAGGAGGATATGAAAGTGATGATGATGATGATGAT 960
QY 127 PheLeuValLeuIleMetArgIleValLeuGlnLeuAlaIysMetAsnLeuMet----- 144
Db 961 AACGCCGAACCTTTCTTC-----TTAGACATTCATAATATTATTCATGTTATGCGGAA 1011
QY 145 AspIleThrIysIlePheSerLeuLeuGlnProAspIysGluGlu----- 159
Db 1012 TCTTTAAAAAAGTAGGACATGTTTATCTCTATGTAGAGAAATCTCATTTGTTGTCC 1071
QY 160 -----GluAspThrAspThrGluGlu-----LysGlnAlaLeuAsnGlnAlaValTyr 175
Db 1072 AGTTTGGAGTCTACTCTATTGGTTAGAACATATCAAGCTCGTTTGTGACAGGACCATCAA 1131
```



```
; NAME/KEY: modified_base
; LOCATION: (212)
; OTHER INFORMATION: Where n is a, c, g or t
US-09-221-298-28

Alignment Scores:
Pred. No.:      1.45e-20      Length:      401
Score:          244.50       Matches:     46
Percent Similarity: 62.26%    Conservative: 20
Best Local Similarity: 43.40% Mismatches:    39
Query Match:      10.44%      Indels:      1
DB:              3           Gaps:        1

US-09-941-831A-20 (1-449) x US-09-221-298-28 (1-401)

QY   5 GlyTTPLyLeuIleAspProLysSerPheGlyArgMetGlyIleProAsnArgAsn 24
Db   71 GGATGGACAGTTTACATCCAGTCGAAGAATACAGGAGGCCTTGCCCAATCACCAT 130

QY   25 TTPThrILEThrASPAlaAsnARGAsnTYRGLULLECYSSserThRTyRPrOPROgluile 444
Db   131 TCGAGAATAACTTTTTATTAAATGAAGTGCTATCAGCTCTGCACACTTACCCTGTCTTTTG 190

QY   45 ValValProLYSserValThrLeuGlyThrValValGlySerSerIysPheArgSerLys 64
Db   191 GTGGTTCCGTATCGTCGCCCTCANATGATGACCTCGGAGAGTTGCAACTTTTAGGTCCCGA 250

QY   65 GluArqValProValLeuSerTYRLeuTYRlYSGLUAsnAsnAlaAlaIleCYsARgCyS 84
Db   251 ATNCGAATTCAGTGTCTGTCATGGATTCCACGAAATAAGACGGTCATTGTGCGGTGC 310

QY   85 SerGlnPROLeuSERglyPheTyR---ThrArgCySValaspGluLeuLeuGlu 103
Db   311 AGTCAGCCTCTTGTGGTATGATGGGAAACGAAATAAGATGATGAGAAATATCTCGAT 370

QY   104 AlaIleSerGlnThrAsn 109
Db   371 GTTATCAGGAGACTAAT 388

RESULT 5
US-09-302-769-26
; Sequence 26, Application US/09302769
; Patent No. 6323317
; GENERAL INFORMATION:
; APPLICANT: HILTON, Douglas J
; APPLICANT: ALEXANDER, Warren S
; APPLICANT: VINEY, Elizabeth M
; APPLICANT: WILLSON, Tracey A
; APPLICANT: RICHARDSON, Rachael T
; APPLICANT: STARR, Robyn
; APPLICANT: NICHOLSON, Sandra E
; APPLICANT: METCALF, Donald
; APPLICANT: NICOLA, Nicos A
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
; FILE REFERENCE: 10976Z
; CURRENT APPLICATION NUMBER: US/09/302,769
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 08/962,560
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 419
; TYPE: DNA
; ORGANISM: Human
; US-09-302-769-26

Alignment Scores:
Pred. No.:      3.45e-19      Length:      419
Score:          233.50       Matches:     49
Percent Similarity: 65.35%    Conservative: 17
Best Local Similarity: 48.51% Mismatches:    32
Query Match:      9.97%      Indels:      3
```

Db 112 GAACGAGTCAAGAGGCTGTGAGCGCTGAGAAGGTGAACAGCGCGGACACGCGGCGGCGAGG 171
Qy 204 ProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeu 223
Db 172 AAATCCACCCGCTGACCTCGCGGAGGTTTGGCGGAAAGACGAGTAGTTGAATATTG 231
Qy 224 LeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPheThr 243
Db 232 CTTCAAGATGGTCAAAATGCCAAGCAGCTGATGATGGGGCTTATTCTCTTCATAAT 291
Qy 244 AlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerPro 263
Db 292 GCATGCTCTTTGGTCAATGCTGAAGTAGTCAATCTCTTTTGGCGAATGCGACATGCGACACCC 351
Qy 264 GlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaVal 283
Db 352 ATGTCGAGATAATGGAAATTATATCTCTCTCATGAAGCTGCAATTAAGAAAGATT 411
Qy 284 AlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeuPro 303
Db 412 GATGTTTGATCTGCTGTGTACAGCATGGAGCTGAGCCCAACCATCCGAATACAGATGA 471
Qy 304 ValTrpAlaSerAsnIleAla----- 310
Db 472 AGGACAGCATTTGATTTAGCAGATCCATCTGCCAAAGCAGTGCTTACTGCTGTAATATAAG 531
Qy 310 ----- 310
Db 532 AAAGATGAACCTTTAGAAAGTCCAGGAGTGGCAATGAAGAAATAATGATGCTCTACTC 591
Qy 311 -----SerCys----- 310
Db 592 ACACCATTAATGCTCACTGCCAGCAGTGGATGGCAGAAAGTCACTCACTTACATTG 651
Qy 319 AlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspPro 338
Db 652 GCAGCAGGATATAACAGAGTAAGATTGTACAGCTGTACTGTCAACATGAGCT----- 705
Qy 339 AspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeu 358
Db 706 GATGTCATGCTAAAGATAAAGGTGATCTG-----GTACCA----- 741
Qy 359 GluIleCysLeuHisHisAsnCys-----GluProGlyTyrIleGlnLeuLeuLeu 375
Db 742 -----TTACACATGCTGCTTCTTATGCTCATATGAAGTAAGTAAGTAAGTAAGT 792

RESULT 7
US-09-350-982C-3

; Sequence 3, Application US/09350982C
; Patent No. 6455290
; GENERAL INFORMATION:
; APPLICANT: Berthelsen, Jens
; APPLICANT: Toma, Salvatore
; APPLICANT: Isacchi, Antonella
; TITLE OF INVENTION: Tankyrase Homolog Protein (THP), Nucleic Acids, and Methods Rela
; TITLE OF INVENTION: Same
; FILE REFERENCE: PHRM-0043
; CURRENT APPLICATION NUMBER: US/09/350,982C
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; TYPE: DNA
; LENGTH: 4512
; ORGANISM: Artificial
; FEATURE:
; NAME/KEY: PCR Primers
; NAME/KEY: misc feature
; LOCATION: (1124)...(1124)

; OTHER INFORMATION: n is any nucleic acid
; NAME/KEY: misc feature
; LOCATION: (2672)...(2672)
; OTHER INFORMATION: n is any nucleic acid
US-09-350-982C-3

Alignment Scores:
Pred. No.: 1,21e-15 Length: 4512
Score: 219.00 Matches: 74
Percent Similarity: 40.98% Conservative: 26
Best Local Similarity: 30.33% Mismatches: 78
Query Match: 9.35% Indels: 66
DB: 4 Gaps: 7

US-09-941-831A-20 (1-449) x US-09-350-982C-3 (1-4512)

Qy 189 GluArgTyrLysArgPhe-----IleAsnSerArgSerGlyTrpGlyVal 203
Db 134 GAACGAGTCAAGAGGCTGTGACGCTGAGAAGGTGAACAGCGCGGACACGCGGCGGCGAGG 193
Qy 204 ProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeu 223
Db 194 AAATCCACCCGCTGACCTCGCGGAGGTTTGGCGGAAAGACGAGTAGTTGAATATTG 253
Qy 224 LeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGlnThrProLeuPheThr 243
Db 254 CTTCAAGATGGTCAAAATGCCAAGCAGCTGATGATGGGGCTTATTCTCTTCATAAT 313
Qy 244 AlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerPro 263
Db 314 GCATGCTCTTTGGTCAATGCTGAAGTAGTCAATCTCTTTTGGCGAATGCGACATGCGACACCC 373
Qy 264 GlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaVal 283
Db 374 ATGTCGAGATAATGGAAATTATATCTCTCTCATGAAGCTGCAATTAAGAAAGATT 433
Qy 284 AlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnValLysAlaLysLeuPro 303
Db 434 GATGTTTGATCTGCTGTGTACAGCATGGAGCTGAGCCCAACCATCCGAATACAGATGA 493
Qy 304 ValTrpAlaSerAsnIleAla----- 310
Db 494 AGGACAGCATTTGATTTAGCAGATCCATCTGCCAAAGCAGTGCTTACTGCTGTAATATAAG 553
Qy 310 ----- 310
Db 554 AAAGATGAACCTTTAGAAAGTCCAGGAGTGGCAATGAAGAAATAATGATGCTCTACTC 613
Qy 311 -----SerCys-----SerGlyProLeuTyrLeu 318
Db 614 ACACCATTAATGCTCACTGCCAGCAGTGGATGGCAGAAAGTCACTCACTTACATTG 673
Qy 319 AlaAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAspPro 338
Db 674 GCAGCAGGATATAACAGAGTAAGATTGTACAGCTGTACTGTCAACATGAGCT----- 727
Qy 339 AspTyrAsnCysThrAspGlnGlyLeuLeuAlaArgValProArgProArgThrLeuLeu 358
Db 728 GATGTCATGCTAAAGATAAAGGTGATCTG-----GTACCA----- 763
Qy 359 GluIleCysLeuHisHisAsnCys-----GluProGlyTyrIleGlnLeuLeuLeu 375
Db 764 -----TTACACATGCTGCTTCTTATGCTCATATGAAGTAAGTAAGTAAGTAAGT 814
Qy 376 AspPheGlyAla 379
Db 815 AAGCATGGTGCC 826

RESULT 8

US-09-196-387-7
; Sequence 7, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:

; APPLICANT: de Lange, Titia
 ; APPLICANT: Smith, Susan
 ; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
 ; TITLE OF INVENTION: OF USE THEREOF
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Klauber & Jackson
 ; STREET: 411 Hackensack Avenue, 4th Floor
 ; CITY: Hackensack
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07601
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/196,387
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 09/095,225
 ; FILING DATE: June 10, 1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jackson Esq., David A.
 ; REGISTRATION NUMBER: 26,742
 ; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201-487-5800
 ; TELEFAX: 201-343-1684
 ; TELEX: 133521

; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4491 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; HYPOTHETICAL: NO
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 6..2027
 ; US-09-196-387-7

Alignment Scores:
 Pred. No.: 5,63e-15 Length: 4491
 Score: 213.50 Matches: 102
 Percent Similarity: 37.50% Conservative: 42
 Best Local Similarity: 26.56% Mismatches: 143
 Query Match: 9.11% Indels: 97
 DB: 3 Gaps: 15

US-09-941-831A-20 (1-449) x US-09-196-387-7 (1-4491)

QY 136 LeuGlnLeuAlaLysMetAsnLeu-----MetAspIleThrLys 148
 Db 2205 TTGCAGCTCTCAAAATGTGAATGTAGAGACTTAGAGGGCGCGCAATTCAGGCGCTTACA 2264
 QY 149 IlePheSerLeuLeuGlnProAspLysGluGluAspThrAspThrGluGluLysGln 168
 Db 2265 CTTCCGAGCAGGCTACACAGAGAGTACACCTATACGTACTTCTGATGTTGATTATCGA--- 2321
 QY 169 AlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeuLeuArgGln 188
 Db 2322 ---CTTTAGAGCATCTAAAGCTGAGACTTGAACACTGTGAACACTTTGCAGCTCT 2378
 QY 189 GluArgTyrLysArgPheIleAsnSerArgSerGlyTyrPValProGlyThrProLeu 208
 Db 2379 CAAAAT-----GTGAATTGAGAGACTTAGAGGCGCGCATTCACGCCCTTA 2426
 QY 209 ArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAlaHisGlyAla 228

Db 2427 CACTTCGCAGCAGGCTACAAACCCGCTGTCTGTGTAGACTACCTGCTACACACCGGTGCC 2486
 QY 229 AspValAspSerLeuAspValLysAlaGlnThrProLeuPheThrAlaValSerHisGly 248
 Db 2487 GATGTCCATGCCAAGACAAGGGTGGTGGCCCTTCATAATGCCTGTTCTATATGGA 2546
 QY 249 HisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerProGlyGlySerIleTyr 268
 Db 2547 CACTATGAGTGGCTGAGCTTTTAGTAAGGCATGGGGCTTCTGTCAATGTGGGGACTTA 2606
 QY 269 AsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaValAlaIleLeuGlnGlu 288
 Db 2607 TGGAAATTTACCCCTCTCCATGAAGCAGCAGCTAAAGGAAGTATGAATCTGCAAGCTC 2666
 QY 289 LeuLeuAspHisGlyAlaGluAlaAsnValLys----- 299
 Db 2667 CTTTAAACATGGAGCAGATCCAACATAAAGAACAGAGATGGAATACACCTTTGGAT 2726
 QY 299 ----- 299
 Db 2727 TTGGTAAAGGAAGGAGACACAGATATTAGGACTTACTGAAAGGGGATGCTGCTTTGTTG 2786
 QY 300 -----AlaLysLeuProValTrpAla-----SerAsnIle 309
 Db 2787 GATGTGCCAAGAAGGGCTGCCTGGCAAGAGTGCAGAACTCTGTACCCAGAGAAATATC 2846
 QY 310 AlaSerCys-----SerGlyProLeuTyrLeuAlaAlaValTyr 322
 Db 2847 ---AACTGCAGACACCCAGGCGCAGAAATTCACCCCTCTGCACCTGCAGCAGGCTAT 2903
 QY 323 GlyHisLeuAspCysPheArgLeuLeuLeuHisGlyAlaAsp----- 337
 Db 2904 AATAACCTGGAAGTAGCTGAATATCTTCTAGAGCATGGAGCTGATGTTAATGCCAGGAC 2963
 QY 338 -----ProAspTyrAsnCysThrAspGlnGly-----LeuLeuAla 349
 Db 2964 AAGGGTGGTTTAATTCCTCTTCAATATGCGGCATCTTATGGGCATGTTGACATAGCGCT 3023
 QY 350 ArgValProArgProArgThrLeuLeuGluIleCys-----Leu 362
 Db 3024 TTATTGATAAAATACACACGCTGTGTAATGCAACAGATAAGTGGCGGTTTACTCCCTC 3083
 QY 363 HisHisAsnCysGluProGluTyrIleGln-----LeuLeuIleAspPheGlyAla 379
 Db 3084 CATGAAGCAGCCAGAAAGGAGGAGCGAGCTGTGCGCCCTCTCTCTAGCGCATGTTGCA 3143
 QY 380 AsnIleTyrLeuProSerLeuSer-----LeuAspLeuThrSer 392
 Db 3144 GAC-----CCACCATGAAGAACCCAGGAAGCCAGACGCTCTGGATCTGGCAACA 3194
 QY 393 GlnAspAspLysGlyIleAlaLeuLeuGlnAla-----ArgAlaThrProArg 409
 Db 3195 GCTGACGATATCAGA---GCTTTGCTGATAGATGCATGCCGCCAGAGGCTTACTTACC 3251
 QY 410 SerLeuLeuSerGlnValArgValLeuValArgAlaLeuCysGlnAlaGlyGlnPro 429
 Db 3252 TGTTTAAACCTCAGGCTACTGTAGTGTAGTGTCTCTGTATCTCACCAGCATCCACCCCC 3311
 QY 430 GlnAlaIleAsn 433
 Db 3312 TCCTGCCCTCTCG 3323

RESULT 9

US-09-841-835-7
 ; Sequence 7, Application US/09841835
 ; Patent No. 6506587
 ; GENERAL INFORMATION:
 ; APPLICANT: de Lange, Titia
 ; APPLICANT: Smith, Susan
 ; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
 ; TITLE OF INVENTION: OF USE THEREOF
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/841,835
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: 09/196,387
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4491 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 6..2027
US-09-841-835-7

Alignment Scores:
Pred. No.: 5.63e-15 Length: 4491
Score: 213.50 Matches: 102
Percent Similarity: 37.50% Conservative: 42
Best Local Similarity: 26.56% Mismatches: 143
Query Match: 9.11% Indels: 97
DB: 4 Gaps: 15

US-09-941-831A-20 (1-449) x US-09-841-835-7 (1-4491)

QY 136 LeuGlnLeuAlaLysMetAsnLeu-----MetAspIleThrLys 148
DB 2205 TTGCAGCTCTCAAAATGTGNAATTTGTAGAGCTTAGAGGGCGCGCATTCACGCCCTTACA 2264

QY 149 IlePheSerLeuLeuGlnProAspLysGluGluAspThrAspThrGluGluLysGln 168
DB 2265 CTTGCGAGCAGGCTACAACAGAGTACACCTATACGTACTCTCATGTTGATATCA--- 2321

QY 169 AlaLeuAsnGlnAlaValTyrAspAsnAspSerTyrThrLeuAspGlnLeuLeuArgGln 188
DB 2322 ---CTCTTAGAGGCATCTAAAGCTGGAGACTTGGAAACTGTGAAGCAACTTTCAGCTCT 2378

QY 189 GluArgTyrLysArgPheIleAsnSerArgSerGlyTrpGlyValProGlyThrProLeu 208
DB 2379 CAAAT-----GTGAATTTGTAGAGACTTAGAGGGCGCGCATTCACGCCCTTA 2426

QY 209 ArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAlaHisGlyAla 228
DB 2427 CACTTCGACAGGCTACACCGCGTCTGTTGTAGAGTACCTGTCTACACAGCGTGCC 2486

QY 229 AspValAspSerLeuLeuValLysAlaGlnThrProLeuPheThrAlaValSerHisGly 248
DB 2487 GATGTCATGCCAAAGACAGAGGGTGGCTTGGTCCCTTCATTAATGCCTGTTTCATATGGA 2546

QY 249 HisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerProGlyGlySerIleTyr 268
DB 2547 CACTATGAGGTGGCTGAGCTTTTAGTAGGCGATGGGCTTCTGTCAATGTGGCGACTTA 2606

QY 269 AsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaValAlaIleLeuGlnGlu 288
DB 2607 TGGAAATTTACCCCTCTCCATGAAGCAGCAGCTAAAGGAAGTATGAAATTCGACGCTC 2666

QY 289 LeuLeuAspHisGlyAlaGluAlaAsnValLys----- 299
DB 2667 CTTTTAAACATGGAGCAGATCCAACTAAAAGAACAGAGATGGAAATACACCTTTGGAT 2726

QY 299 ----- 299
DB 2727 TTGGTAAAGGAAGGAGACACAGATATTCAGGACTTACTGAAAGGGGATGCTGCTTTGTTG 2786

QY 300 -----AlaLysLeuProValTrpAla-----SerAsnIle 309
DB 2787 GATGCTGCCAAGAGGGGCTGCTGCGCAAGAGTGCAGAGCTCTGTACCCAGAGATATC 2846

QY 310 AlaSerCys-----SerGlyProLeuTyrLeuAlaAlaValTyr 322
DB 2847 ----AACTGCAGAGACACCCAGGCGCAAAATTCACCCCTCTGCACCTGCAGAGGCTAT 2903

QY 323 GlyHisLeuAspCysPheArgLeuLeuHisGlyAlaAsp----- 337
DB 2904 AATAACCTGGAAGTAGCTGAATATCTCTAGAGCATGGAGCTGATGTTAATGCCAGGAC 2963

QY 338 -----ProAspTyrAsnCysThrAspGlnGly-----LeuLeuAla 349
DB 2964 AAGGGTGGTTAAATTCCTCTCATTAATGCGGCATCTTATGGGCATGTTGACATAGCGCT 3023

QY 350 ArgValProArgProArgThrLeuLeuGluIleCys-----Leu 362
DB 3024 TTATTGATAAAATACACACCGTGTGTAATGCAACAGATAGTGGCGGTTTACTCCCCCTC 3083

QY 363 HisHisAsnCysGluProGluTyrIleGln-----LeuLeuIleAspPheGlyAla 379
DB 3084 CATGAAGCAGCCAGAGAGGAGGAGCGAGCTGTGGCCCTCTCTCTAGCGCATGTTGCA 3143

QY 380 AsnIleTyrLeuProSerLeuSer-----LeuAspLeuThrSer 392
DB 3144 GAC-----CCACCATGAAGAACACAGGAAGCGCAGACGCTCTGTGATCTGGCAACA 3194

QY 393 GlnAspAspLysGlyIleAlaLeuLeuGlnAla-----ArgAlaThrProArg 409
DB 3195 GCTGACGATATCAGA---GCTTTGCTGATAGATGCCATGCCCCAGAGCGCTTACCTACC 3251

QY 410 SerLeuLeuSerGlnValArgLeuValValArgAlaLeuCysGlnAlaGlyGlnPro 429
DB 3252 TGTTTTAAACCTCAGGCTACTGTAGTGAGTGCTCTCTGATCTCACCAGCATCCACCCC 3311

QY 430 GlnAlaIleAsn 433
DB 3312 TCCTGCTCTCG 3323

RESULT 10
US-09-196-387-9
; Sequence 9, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601


```

; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4657 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6..2855
; US-09-841-835-9

Alignment Scores:
Pred. No.: 2,12e-14 Length: 4657
Score: 209.00 Matches: 93
Percent Similarity: 36.28% Conservative: 30
Best Local Similarity: 27.43% Mismatches: 122
Query Match: 8.92% Indels: 94
DB: 4 Gaps: 11

US-09-941-831A-20 (1-449) x US-09-841-835-9 (1-4657)

QY 139 AlaLysMetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGlu 158
DB 1887 GCACAGATGGCAATGAAGCAGTGCAGAGATCTCAGTGAGAGTACACTATACGT--- 1943
QY 159 GluGluAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAsp 178
DB 1944 ACTTCGTGATGTTGAT-----TATCGACTCTTAGAGGCATCTAAAGCTGGAGAC 1991
QY 179 SerTyrThrLeuAspGlnLeuArgGlnGluArgTyrLysArgPheIleAsnSerArg 198
DB 1992 TTGGAAACTGTGAAGCAACTTTGCAGCTCTCAAAAT-----GTGAATTGTAGA 2039
QY 199 SerGlyTyrGlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSer 218
DB 2040 GACTTAGAGGGCGGCGATTCACAGCCCTTACACTTCGAGCAGGCTACACCGCGTGT 2099
QY 219 CysLeuGlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGln 238
DB 2100 GTTGTAGAGTACTGTGTACACACGCGTCCGATGTCATGCCAAAGACAAAGGCTGG 2159
QY 239 ThrProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGlu 258
DB 2160 GTGCCCTTTTCATAATGCCCTTTCATATGGACACTATGAGGTGGCTGAGCTTTTGTAG 2219
QY 259 AlaGlyAlaSerProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAla 278
DB 2220 CATGGGGCTTCTGTCATGTGGGGGACTTATGGAAATTTACCTCTCCATGAAGCAGCA 2279
QY 279 ArgAspGlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnVal 298
DB 2280 GCTAAAGGAAAGTATCAATCTGCAAGCTCTTTTAAACATGGACGACATCAACTAAA 2339
QY 299 Lys----- 299
DB 2340 AAGAACAGAGATGGAAATACCTTTGGATTTGGTAAAGGAGGAGACACAGATATTCAG 2399
QY 300 -----AlaLysLeuProValTyrAla--- 306
DB 2400 GACTTACTGAAAGGGGATGCTGTTTGGATGTGCCCAAGAGGGCTGCTGGCAAGA 2459
QY 307 -----SerAsnIleAlaSerCys----- 312
DB 2460 GTGCAGAGCTCTGTACCCACAGAGATATC---AAGTGCAGACACCCAGGGCAGAAAT 2516
QY 313 SerGlyProLeuTyrLeuAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeu 332
DB 2517 TCAACCCCTCTGCACCTGGCAGCAGCTGATATAACCTGGAAGTAGCTGAATATCTCTA 2576

```

```

QY 333 LeuHisGlyAlaAsp-----ProAspTyrAsnCys 342
DB 2577 GAGCATGGAGCTGATGTTAATGCCAGCAGCAAGGCTGTTAATTCCTCTTCATAATGCG 2636
QY 343 ThrAspGln---GlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluLeuCys 361
DB 2637 GCATCTTATGGGGCTGCTGCTGCAAGAGTGCAGAAG----- 2672
QY 362 LeuHisAsnCysGluProGluTyrIleGlnLeuLeuLeuAspPheGlyAlaAsnIle 381
DB 2673 -----CTCTGTACCCAGAGATATCACTGCAGACACCCAGGGCAGAAATTC 2723
QY 382 -----TyrLeuProSer 385
DB 2724 ACCCTCTGCACCTGGCAGCAGGCTAATAACCTGGAAGTAGCTGAATATCTTCTAGAG 2783
QY 386 LeuSerLeuAspLeuThrSerGlnAspLysGlyIleAlaLeuLeuLeuGlnAla 404
DB 2784 CATGGAGCTGATGTTAATGCCAGCAGCAAGGCTGTTAATTCCTCTTCATAATGCG 2840

RESULT 12
US-09-196-387-1
; Sequence 1, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRP1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4134 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; US-09-196-387-1

Alignment Scores:
Pred. No.: 1,99e-14 Length: 4134
Score: 208.50 Matches: 100
Percent Similarity: 37.70% Conservative: 41
Best Local Similarity: 26.74% Mismatches: 140
Query Match: 8.90% Indels: 93

```

DB: 3 Gaps: 15

US-09-941-831A-20 (1-449) x US-09-196-387-1 (1-4134)

Qy 139 AlalyeMetAsnLeuMetAspIleThrLysIlePheSerLeuLeuGlnProAspLysGlu 158
 Db 1887 GCACAGATGGCAATGAGCAGGAGATCTCTGAGTGAGAGTACACCTATAGCT--- 1943
 Qy 159 GluGluAspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAspAsnAsp 178
 Db 1944 ACTTCTGATGTTGAT-----TATCGACCTTAGAGCGCATCTAAAGCTGGAGAC 1991
 Qy 179 SerTyrThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsnSerArg 198
 Db 1992 TTGGAACCTGTGAAGCAACTTTCAGCTCTCAAAAT-----GTGAATTTGAGA 2039
 Qy 199 SerGlyTrpGlyValProGlyThrProLeuArgLeuAlaAlaSerTyrGlyHisLeuSer 218
 Db 2040 GACTTAGAGGCGCGCATTCACGCGCTTACACTTCGCAGCAGGCTACACCGCGTGTCT 2099
 Qy 219 CysLeuGlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspVallyAlaGln 238
 Db 2100 GTTGTAGACTGCTGACACAGCGTGCAGATGTCATGCCAAGACAAGGGTGGCTTG 2159
 Qy 239 ThrProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGlu 258
 Db 2160 GTGCCCCCTTCATAATGCCTGTTTCATATGACACTATGAGTGGCTGAGCTTTTAGTAGG 2219
 Qy 259 AlaGlyAlaSerProGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAla 278
 Db 2220 CATGGGCGCTCTGTCATATGGCGGACTTATGGAATTTACCCCTCTCCATGAAGCAGCA 2279
 Qy 279 ArgAspGlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnVal 298
 Db 2280 GCTAAGGAAGATGAAATCTGCAAGCTCTTTTAAACATGGAGCAGATCCCACTAAA 2339
 Qy 299 Lys-----SerAsnIleAlaSerCys----- 299
 Db 2340 AAGACAGAGATGAAATATACACTTTGGATTTGGTAAAGGAAGAGACACAGATATTACAG 2399
 Qy 300 -----AlaLysLeuProValTrpAla--- 306
 Db 2400 GACTTACTGAAGGGGATGCTGCTTTGTTGGATGCTGCCAAGAAAGGGCTGCCTGGCAAGA 2459
 Qy 307 -----SerAsnIleAlaSerCys----- 312
 Db 2460 GTGCAGAGCTCTGTACCCAGAGATATC---ACTGCAGAGACACCCAGGGCAGAAAT 2516
 Qy 313 SerGlyProLeuTyrLeuAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeu 332
 Db 2517 TCAACCCCTCTGCACCTGCAGCAGCTATATAACCTGGAAGTAGCTGATATCTCTTA 2576
 Qy 333 LeuHisGlyAlaAsp-----ProAspTyrAsnCys 342
 Db 2577 GAGCATGGAGCTGATGTTAATGCCAGGACAGGGTGGTTTAAATTCCTCTTCATAATGCG 2636
 Qy 343 ThrAspGlnGly-----LeuLeuAlaArgValProArgProArgThrLeuLeuGlu 359
 Db 2637 GCATCTATGGGCATGTTGACATAGCGGCTTTATGATAAATACACACCTGTGTAAT 2696
 Qy 360 IleCys-----LeuHisHisAsnCysGluProGluTyrIleGln 372
 Db 2697 GCAACAGATAAGTGGCGCTTACTCCCTCCATGAAGCAGCCGCAAGGAAGGAGCCGAG 2756
 Qy 373 -----LeuLeuIleAspPheGlyAlaAlaSerIleTyrLeuProSerLeuSer----- 387
 Db 2757 CTGTGCGCCCTCTCTCTAGCGATGGTGACAG-----CCACCATGAAGAACCCAG 2807
 Qy 388 -----LeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeuLeuLeu 402
 Db 2808 GAAGGCCACACGCTCTGGATCTGGACACAGCTGACGATATCAGA---GCTTTCTCTGATA 2864
 Qy 403 GlnAla-----ArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValVal 419

```
Db 1992 TTGGAACCTGTGAAGCAACTTTTGACGCTCTCAAAAT- - - - -GTGAATGTGACA 2039
QY 199 SerGlyTyrProGlyValProGlyThrProLeuArgLeuAlaSerTyrGlyHisLeuSer 218
Db 2040 GACTTAGAGGGCGGCGGATTCACCGCCCTTACACTTCGCGACGGCTACCAACGCGGTGTCT 2099
QY 219 CysLeuGlnValLeuLeuAlaHisGlyAlaAspValAspSerLeuAspValLysAlaGln 238
Db 2100 GTTGTAGAGTACTGTACACACGGTGGCGATGTCATGCCAAGACAGGGTGGCTTG 2159
QY 239 ThrProLeuPheThrAlaValSerHisGlyHisLeuAspCysValArgValLeuLeuGlu 258
Db 2160 GTGCCCTTCAATAATGCTTTCATATGACACTATGAGTGGCTGAGCTTTTAGTAAGG 2219
QY 259 AlaGlyAlaSerProGlySerGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAla 278
Db 2220 CATGGGGCTTCTGCAATGTGGCGGACTATGGAATTTACCCCTTCCATGAAGCAGCA 2279
QY 279 ArgAspGlyAlaValAlaIleLeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnVal 298
Db 2280 GCTAAAGGAAGTATGAAATCTCAAGCTCTTTTAAACATGGACGACATCCAACTAA 2339
QY 299 Lys- - - - - 299
Db 2340 AAGAACAGAGATGAAATACACCTTTTGGATTGTTAAAGGAAGGACACAGATATTACG 2399
QY 300 - - - - -AlaLysLeuProValTrpAla- - - - - 306
Db 2400 GACTTACTGAAAGGGGATGCTGTTTGGTGGTCTGCCAAGAGGGCTGCGTGGCAAGA 2459
QY 307 - - - - -SerAsnIleAlaSerCys- - - - - 312
Db 2460 GTCCAGAGCTGTGTACCCAGAGATATC- - - - -AATGAGAGACACCCAGGCGCAGAAAT 2516
QY 313 SerGlyProLeuTyrLeuAlaValTyrGlyHisLeuAspCysPheArgLeuLeuLeu 332
Db 2517 TCAACCCCTCTGCACCTGGCAGCAGGCTATAATACTGGAAGTAGCTGAATATCTTCTA 2576
QY 333 LeuHisGlyAlaAsp- - - - -ProAspTyrAsnCys 342
Db 2577 GAGCATGGAGCTGATGTTAATGCCAGGACAGGGTGGTTTAAATCTCTCTCATATGCG 2636
QY 343 ThrAspGlnGly- - - - -LeuLeuAlaArgValProArgProArgProArgLeuLeuGlu 359
Db 2637 GCATCTTATGGCATGTTGACATAGCGGCTTATTGATAAAATACAAACGCTGTGTAAT 2696
QY 360 IleCys- - - - -LeuHisAsnCysGluProGluTyrIleGln 372
Db 2697 GCAACAGATAAGTGGGCTTTACTCCCTCCATGAAGCAGCCAGAAAGGAGGACGACG 2756
QY 373 - - - - -LeuLeuIleAspPheGlyAlaAsnIleTyrLeuProSerLeuSer- - - - - 387
Db 2757 CTGTGGGCCCTCTCTCTAGCGCATGGTGGCAGAC- - - - -CCCACCATGAAGAACGAC 2807
QY 388 - - - - -LeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeuLeuLeu 402
Db 2808 GAAGGCCAGAGCGCTCTGGATCTGGCAACAGCTGACGATATCAGA- - - - -GCTTTGCTGATA 2864
QY 403 GlnAla- - - - -ArgAlaThrProArgSerLeuLeuSerGlnValArgLeuValVal 419
Db 2865 GATGCCATGCCCCAGAGGCTTACCTACTGTTTAAACCTCAGGCTACTGTAGTGTAGT 2924
QY 420 ArgArgAlaLeuCysGlnAlaGlyGlnProGlnAlaIleAsn 433
Db 2925 GCCTCTCTGTATCTCACCAGCATCCACCCCTCTCTGCTCTCG 2966
```

RESULT 14

```
US-09-620-312D-1101
; Sequence 1101, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
```

```
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Wang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yungqing
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: John Tillinghast
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. 6569662el Nucleic Acids and
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pt_FL_genes Version 1.0
SEQ ID NO 1101
LENGTH: 1712
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (551)..(1672)
US-09-620-312D-1101
```

```
Alignment Scores:
Pred. No.: 6,68e-15 Length: 1712
Score: 207.00 Matches: 92
Percent Similarity: 42.90% Conservative: 44
Best Local Similarity: 29.02% Mismatches: 111
Query Match: 8.83% Indels: 70
DB: 4 Gaps: 14
```

US-09-941-831A-20 (1-449) x US-09-620-312D-1101 (1-1712)

```
QY 157 LysGluGluGluAspThrAspThrGluGluLysGlnAlaLeuAsnGlnAlaValTyrAsp 176
Db 208 CGCCAGAGTCAAGACTGGGGCAAGAGTGACGAGAGGCTGCTA- - -CAAGCCGTGGAAGAC 264
QY 177 AsnAspSerTyrThrLeuAspGlnLeuLeuArgGlnGluArgTyrLysArgPheIleAsn 196
Db 265 AACGATGACCTCGGGTGGCCCGCTCATCGCCCGCAAG- - - - - 303
QY 197 SerArgSerGlyTrpGlyValProGlyThrProLeu- - - - - 208
Db 304 - - - - -GGCTGTGTGCCAGCAAGCTAGACCCCGAGGGCAAGTCCGCGTTC 348
QY 209 ArgLeuAlaAlaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAlaHisGlyAla 228
Db 349 CACCTGGGGCCATCGGGGTGGCGGACAGTCTGTGGAGGTGATGATAGCTCATGGCAGC 408
QY 229 AspValAspSerLeuAspValLysAlaGlnThrProLeuPheThrAlaValSerHisGly 248
Db 409 AATGTATCATGAGCGGACGAGCGGCGAGTTACAATGCCCTCCACCTGGCCGCAAAATACGG 468
QY 249 HisLeuAspCysValArgValLeuLeuGluAlaGly-Ala- - - - -Se 262
Db 469 CACCCACAGTGTCTGAAGCAACTACTGACGGCTTCTGCTGTGTGGACCTCTGTGGACAGC 528
QY 262 rProGlyGlySerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAl 282
Db 529 AGCGGGTGGAGTGCCTTACACCATGCGGGGTGGATTGCTTCCCTAGCAGCGGCTGTGCTG 598
```

```
QY 282 aValAlaIleLeuGlnGluLeuAspHisGlyAlaGluAlaAsnVallylsAlaLysLe 302
Db 589 TCCTCTCTGCTCAGAGGTGCTGCTCTCTTAAGGCACATCTAAACCCCAAGATCGG-- 646
QY 302 uProValTrpAlaSerAsnIleAlaSerCysGlyProLeuTyrLeuAlaValTy 322
Db 647 -----TCAGGCGCAACACACCTCTCATTTATAGCAGTCAGAT 681
QY 322 rGlyHisLeuAspCysPheArgLeuLeuHisGlyAlaAspProAspTyrAsnCy 342
Db 682 GTGTACACAGACTGTGCGCTCTCTACTGCGAAGGGGCT-----GCCGC 729
QY 342 sThrAspGlnGlyLeuAlaArgValProArgProArgThrLeuLeuGluIleCysLe 362
Db 730 GAACGATCAGACCTGCAAGCGCAGGCG-----GCCCTGATGCTGCGCTGT-- 775
QY 362 uHisAsnCyGluProGluTyrIleGlnLeuLeuLeuAspPheGlyAlaAsnIleTy 382
Db 776 ----GAGGGGGCCAGCGCCGAAACAGTGGAGTCTCTGCTGAGGGGGGAGCCAG----- 826
QY 382 rLeuProSerLeuSerLeuAspLeuThrSerGlnAsp----- 394
Db 827 ----CCGGGATCACC-----GATGCGTGGGGCAGGCGGCTCACTATGGCGCCCTGGC 879
QY 395 ----AspLysGlyIleAlaLeuLeuGln-----AlaArg-AlaThrProArgSerL 411
Db 880 GGGGGACAACTACTCTGCTGCTTCTGCAAGAGGGCGGCCAGCGCCCTCCGCCACCCAG 939
QY 411 euLeuSerGlnValArgLeuValValArg-----ArgAlaLeuCysGlnAla---- 426
Db 940 CGCCCTCACAGAGGATGATTCAGGCGAGGGGCTCATCTCAGAACTCTATGTCCAGCCATGG 999
QY 427 -----GlyGlnProGlnAlaIleAsnGlnLeuAspIleProPro 439
Db 1000 AAAGCAGGGGGGGCCCCCAAGAGCGGAAGCGGCTCCACCTCCCGCCA 1046

RESULT 15
US-09-220-132-30
; Sequence 30, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
; OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 07334-074001
; CURRENT APPLICATION NUMBER: US/09/220,132
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/068,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 14770
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-220-132-30
```

```
Alignment Scores:
Pred. No.: 4,75e-13 Length: 14770
Score: 205.00 Matches: 63
Percent Similarity: 45.42% Conservative: 46
Best Local Similarity: 26.25% Mismatches: 93
Query Match: 8.75% Indels: 38
DB: 4 Gaps: 9
```

US-09-941-831A-20 (1-449) x US-09-220-132-30 (1-14770)

```
QY 206 ThrProLeuArgLeuAlaSerTyrGlyHisLeuSerCysLeuGlnValLeuLeuAla 225
Db 1198 TCTCCATGACATGTCGCGCCACACAAAGGGGATCATTTAAACTGGCTCCAGCTTCTCTCCAG 1257
```

```
QY 226 HisGlyAlaAspValAspSerLeuAspVallylsAlaGlnThrProLeuPheThrAlaVal 245
Db 1258 CATATGTATACCGTGGATGATGTACCATGTACTACCTGCTACCTGCGGTGGC 1317
QY 246 SerHisGlyHisLeuAspCysValArgValLeuLeuGluAlaGlyAlaSerProGlyGly 265
Db 1318 CACTGTGGGCATTAACAAGTTGCAAGGTCTCTTTGGATAAGAAAGCTAACCCCAATGCC 1377
QY 266 SerIleTyrAsnAsnCysSerProValLeuThrAlaAlaArgAspGlyAlaAlaIle 285
Db 1378 AAAGCCCTGAATGGCTTTACCCCTCTTCATATCGCTGCAAGAAGAAATCGAATTAAGTA 1437
QY 286 LeuGlnGluLeuLeuAspHisGlyAlaGluAlaAsnVallylsAlaLysLeuProValTrp 305
Db 1438 ATGGAACTCTCTTGAACACCGT-----GCATCCATCCAAGCT-----GTAACC 1482
QY 306 AlaSerAsnIleAlaSerCysSerGlyProLeuTyrLeuAlaAlaValTyrGlyHisLeu 325
Db 1483 GAGTCGGGCTTACC-----CCAATCCATGTTGCTGCTTCATGGGCATGTA 1530
QY 326 AspCysPheArgLeuLeuLeuHisGlyAlaAspProAspTyrAsnCysThrAspGln 345
Db 1531 AATATTGTATCAACTAATGCATCATGGAGCCTCACCA-----AACACCACCAAT--- 1581
QY 346 GlyLeuLeuAlaArgValProArgProArgThrLeuLeuGluIleCysLeuHisAsn 365
Db 1582 -----GTGAGAGGAGAAACAGCAGCTGCACATGGCAGCTCGCTCCGCG 1623
QY 366 CysGluProGluTyrIleGlnLeuLeuLeuAspPheGlyAlaAsnIleTyrLeuProSer 385
Db 1624 ---CAAGCTGAAGTTGTGCGGTATCTGTTACAAGACGGAGCTCAGGTA----- 1668
QY 386 LeuSerLeuAspLeuThrSerGlnAspAspLysGlyIleAlaLeuLeuGlnAlaArg 405
Db 1669 -----GAAGCTAAAGCTAAGGATGACCAA----- 1692
QY 406 AlaThrProArgSerLeuLeuSerGlnValArgLeu-----ValValArgArgAlaLeu 423
Db 1693 ---ACACCACTCCACATTTACGCCGCTGGGGAAGCAGACATAGTACACAGCTGTTG 1749
QY 424 CysGlnAlaGlyGlnProGlnAlaIleAsnGlnLeuAspIleProProMetLeuIleSer 443
Db 1750 CAGCAAGGGGGCATCTCCAAATGCAGCCCAACTTCTGGGTACACCCCACTTCCCTTTCC 1809
```

Search completed: December 1, 2003, 09:51:26
Job time : 119 secs

RESULT 5

US-09-155-078-2

Sequence 2, Application US/09155078

Patent No. 6312688

GENERAL INFORMATION:

APPLICANT: POUSTKA, Annemarie

APPLICANT: KIOSCHIS, Petra

APPLICANT: LAPORTE, Jocelyn

APPLICANT: HU, Ling Jia

APPLICANT: MANDEL, Jean Louis

APPLICANT: DAHL, Niklas

TITLE OF INVENTION: TYROSINE-PHOSPHATASE-RELATED PROTEIN

FILE REFERENCE: 012627-010

CURRENT APPLICATION NUMBER: US/09/155,078

CURRENT FILING DATE: 1999-03-02

EARLIER APPLICATION NUMBER: PCT/DE97/00592

EARLIER FILING DATE: 1997-03-21

EARLIER APPLICATION NUMBER: DE 196 11 234.6

EARLIER FILING DATE: 1996-03-21

NUMBER OF SEQ ID NOS: 4

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 621

TYPE: PRT

ORGANISM: German-type microorganism & cell cul.

US-09-155-078-2

Query Match

Best Local Similarity 11.3%; Score 264; DB 4; Length 621;

Matches 80; Conservative 46; Mismatches 106; Indels 68; Gaps 12;

QY 5 GWKLDIPISDFGRMGIPNRWTTIDANRNYEICTYPPETVVPKSVTLGTWVGSSKFRSK 64

DB 181 GWTVPNVEYRQGLNHWRTFKVCYELCDTPALLVFPYRASDDDLRRVATFRSR 240

QY 65 ERYVLSYLYKNNAAICRCSQPLSGFY-TRCVDDLELLLEAISQTPGSGQFMVTVTRPK 123

DB 241 NRIPVLSWIHPNKTVIRCSQPLSGMGRKNDKYLVDVIRETNKQISKLTIDYDAPS 300

QY 124 I-----WHFLVLMRIVLQAKMNL--DITKIFSLLOPDKEE----- 159

DB 301 VNAVAKATGGVSDAYNAELFF--LDIHNHVRRESLKKVKDVIYVNVVEESHWLS 357

QY 160 --EDTTEE--KQALNOAVYDNDSYTLQDLRQERYKRFINSRGMGVPGCTPLRLA---- 211

DB 358 SLESTHLEHIKLVLTGAIQVADK-----VSSCKSSVLVHCSGDGMDRTAQLTSLAML 411

QY 212 -----ASYGHLSCLOVLLAHGADVDSLDVKAQTPLFYAVSHGHLDV 253

DB 412 DSFVRSIEGFELLVQKEWISFGHKFASRI--GHG-DKNHTDAD-RSPIFLQ----FIDCV 463

RESULT 6

US-09-350-982C-5

Sequence 5, Application US/09350982C

Patent No. 6455290

GENERAL INFORMATION:

APPLICANT: Berthelsen, Jens

APPLICANT: Toma, Salvatore

APPLICANT: Isacchi, Antonella

TITLE OF INVENTION: Tankyrase Homolog Protein (THP), Nucleic Acids, and Methods Rel

FILE REFERENCE: PHRM-0043

CURRENT APPLICATION NUMBER: US/09/350,982C

CURRENT FILING DATE: 1999-07-09

NUMBER OF SEQ ID NOS: 10

SOFTWARE: PatentIn version 3.0

SEQ ID NO 5

LENGTH: 1166

TYPE: PRT

ORGANISM: Artificial

FEATURE:

NAME/KEY: misc feature

OTHER INFORMATION: Xaa is any amino acid

NAME/KEY: misc feature

LOCATION: (1102)..(1102)

OTHER INFORMATION: n is any nucleic acid

NAME/KEY: misc feature

LOCATION: (2650)..(2650)

OTHER INFORMATION: n is any nucleic acid

US-09-350-982C-5

Query Match

Best Local Similarity 9.3%; Score 219; DB 4; Length 1166;

Matches 74; Conservative 26; Mismatches 78; Indels 66; Gaps 7;

QY 189 ERYKRF-----INSRSGMGVPGTPTLRLLAASYGHLSCLOVLLAHGADVDSLDVKAQTP 243

DB 38 ERVKELVTPPEKVNRSRDTAGRKSTPLHLAAGFGKRVQDVVEYLLQNGANVQARDGGGLPLHN 97

QY 244 AVSHGHLDVVRVLEAGASPGSIYNNCSPLVTARDCGAVAILQELLDHGAENVKAKLP 303

DB 98 ACSFGHAEVNVLLRHGADPNARDNNWYTPHLHBAIKGIDVICVILLOHGAEPTRINTDG 157

QY 304 VNASNIA-----SC-----SGPLYL 318

DB 158 RTALDLADPSAKAVLTGEYKKDELLESARSNGNEEKQWALLTPLNVNCHASDGRKSTPLHL 217

QY 319 AAVYGHLCDFRLLHLHGADPDYNTDQGLLARVPRTTLEICLHNC---BPEYIQLLI 375

DB 218 AAGYNRVKIVQLLOHGA--DVHAKDKGDL--VP-----LHNACSYGHYEYVELLV 264

QY 376 DEGA 379

DB 265 KHGA 268

RESULT 7

US-09-196-387-10

Sequence 10, Application US/09196387

Patent No. 6277613

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

TITLE OF INVENTION: OF USE THEREOF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSER: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th Floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/196,387

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/095,225

FILING DATE: June 10, 1998

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-230 CIP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

TELEX: 133521

THIS PAGE BLANK (USPTO)

Db 347 GGGCCGAGGTGGACCTGGTGGATGTCAAGGGGAGAGCTGCCCTGATGGCTGTAGG 406
Qy 1105 CATGGCCATCTGGAGCTGTGTACGTGTCTTTTGAAGCTGGCTCTCTGGGTGAGC 1164
Db 407 AACGGGCACTGGAGAGAGCACTGAGATCTTTTGGAGGCTGGCTGATCCCAACGGCAGC 466
Qy 1165 ATCTCAACAACCTGTCTTCCCGTGTCAAGCTGGCCCGATGAGTGGCTGTATCCG 1224
Db 467 CGGACCAACCCGAGCACTCTCTGTGTACATGCTCTTGTGTGGAGGAGCAATCTCG 526
Qy 1225 CAGAGCTCTAGACATGTGTGACAGAGGCAAGTCAAGCTAACT-----ACC 1274
Db 527 AAGGCTCTTATCAGGATGAGGGGAGATGTGTATCAACATCATGTGAATTCTGACACC 586
Qy 1275 AGTCTGGGATCAAAATAGCTTCAATGTCT-----GGCCCTCTATTGGCCGCA 1326
Db 587 CGGCCCCCTTTTTCAGGGGGCTTAACCTCTTGGTGTGTCTCTTATACATCAGTGT 646
Qy 1327 GTCTACGGGCACTCTGAGCTGTCTTCCGCTGTCTTTTGTCTCAAGGGGAGAGCCCTGACTAC 1386
Db 647 GCTTACCATTAACCTTCAAGTCTTCAAGGCTGTCTTGTGAGGCTGGGAGCAATCTGACTTC 706
Qy 1387 AACTGCACTGACCAAGGCG 1404
Db 707 AATGCAATGGCCCTGTGTC 724

RESULT 2

US-09-302-769-26
Sequence 26, Application US/09302769

Patent No. 6323317
GENERAL INFORMATION:

APPLICANT: HILTON, Douglas J
APPLICANT: ALEXANDER, Warren S
APPLICANT: VINEY, Elizabeth M
APPLICANT: WILSON, Tracey A
APPLICANT: RICHARDSON, Rachael T
APPLICANT: STARR, Robyn
APPLICANT: NICHOLSON, Sandra E
APPLICANT: METCALF, Donald
APPLICANT: NICOLA, Nicos A
TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
FILE REFERENCE: 109762
CURRENT FILING DATE: 1999-04-30
PRIOR FILING DATE: 08/962,560
NUMBER OF SEQ ID NOS: 50
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 26
LENGTH: 419
TYPE: DNA
ORGANISM: Human
US-09-302-769-26

Query Match 3.8%; Score 71.2; DB 4; Length 419;
Best Local Similarity 55.4%; Pred. No. 2.2e-11;
Matches 164; Conservative 0; Mismatches 123; Indels 9; Gaps 1;

Qy 862 GAGAGACAGGCTCTCAATCAAGAGTGTATGACAAAGTCTCTATCTTTGAGCAAGCTT 921
Db 114 GAGGACAGAGGCTCTCATATGAGCTTACGTGGGAGCTTCAAGACCTTCAGAGACTTA 173
Qy 922 TTGGCCAGAGAGCTTCAAAAGCTTCAATCAAGAGAGAGTGGCTGG-----GGT 972
Db 174 TTGCAAGAGAGAGAGTCAACGAGAGCCCATCAAGAAAGTGTGTGTGTGTGGCTGG 233
Qy 973 GTTCTGGAGACCTTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1032
Db 234 CTCCTTGGAGACCTTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 293
1033 CTCCTTGGAGAGCTGT 1092

Db 294 CTATCCGGAAGGGGCGGAGGTGATCTGTGGAGCTAAAGGAGAGAGCCCTGTAT 353
Qy 1093 ACTGCTGTCACTATGAGCCATCTGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1148
Db 354 GTGGCTGTGTGTAAGGGGAGCTAGAGATGATCAACCAAGATCTCTTCAAGAGTGGCGC 409

RESULT 3

US-09-221-298-28
Sequence 28, Application US/09221298

Patent No. 6284241
GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND DIAGNOSIS
FILE REFERENCE: 210121.471
CURRENT FILING DATE: 1998-12-23
NUMBER OF SEQ ID NOS: 112
SOFTWARE: PasteSeq for Windows Version 3.0
SEQ ID NO 28
LENGTH: 401
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: modified_base
LOCATION: (212)
OTHER INFORMATION: Where n is a, c, g or t
US-09-221-298-28

Query Match 3.7%; Score 69.4; DB 3; Length 401;
Best Local Similarity 52.4%; Pred. No. 7.4e-11;
Matches 176; Conservative 0; Mismatches 157; Indels 3; Gaps 1;

Qy 360 AAAAGAGTGAAGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 419
Db 52 AGAAGATTTAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 111
Qy 420 GGAATATCCCAAGAAATGAGACCTATGACATGATGATGATGATGATGATGATGATG 479
Db 112 GGGCTTCCCAATCAACATGAGAAATTAATTAATTAATTAATTAATTAATTAATTAAT 171
Qy 480 CACCTACCTCTCTGAATTAAGTGTCTTAATCTGTAATCTGTAATCTGTAATCTGTAAT 539
Db 172 CACTTACCTCTCTTTTGT 231
Qy 540 TTCAAGTTCAGAGTAAAGACCTGTCTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 599
Db 232 TGCAACTTTAAGTCCGGAATGAAATTCAGATGCTGTGTGTGTGTGTGTGTGTGTGTGT 291
Qy 600 TGCTGCAATTTGCGCTGT 656
Db 292 GACGCTATGT 351
Qy 657 TGATGAGTCTTGTGTGAGGCGCATTAAGCAAGAA 692
Db 352 TGATGAGAAATATCTCATGTTATCAAGAGACTTA 387

RESULT 4

US-09-155-078-1
Sequence 1, Application US/09155078

Patent No. 6312688
GENERAL INFORMATION:

APPLICANT: POUSTKA, Annemarie
APPLICANT: KIOSCHIS, Petra
APPLICANT: LAPORTE, Jocelyn
APPLICANT: MANDEL, Jean Louis
APPLICANT: DAHL, Niklas
TITLE OF INVENTION: TYROSINE-PHOSPHATASE-RELATED PROTEIN
FILE REFERENCE: 012627-010
CURRENT APPLICATION NUMBER: US/09/155,078

CURRENT FILING DATE: 1999-03-02
; EARLIER APPLICATION NUMBER: PCT/DE97/00592
; EARLIER FILING DATE: 1997-03-21
; EARLIER APPLICATION NUMBER: DE 196 11 234.6
; EARLIER FILING DATE: 1996-03-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3431
; TYPE: DNA
; ORGANISM: German-type microorganism & cell cul.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1863)
US-09-155-078-1

Query Match 3.6%; Score 67.2; DB 4; Length 3431;
Best Local Similarity 52.1%; Pred. No. 1.3e-09;
Matches 175; Conservative 0; Mismatches 158; Indels 3; Gaps 1;
Qy 360 AAAAGAGATGGGAAAGTGGATGGAAGTGAATGACCCCAATATCAGACTTTGGGCGTAT 419
Db 522 AGAAAGTTTAACTGGATGGATGACAGTTTTACAATCCAGTGGGAAGTAATACAGAGGCA 581
Qy 420 GGAATATCCCAACAGAACTGACCATAAACAGATGCCCAACAGAACTATGAGATATGCAG 479
Db 582 GGGCTGCCCATCACCATTGGAGAAATATTTTATTAAATAGTCTATGAGCTCTGTGA 641
Qy 480 CACCTACCTCTCTGAAATAGTGGTTCCTTAATATCTGTACCTTGGGAACGGTGGTGAAG 539
Db 642 CACTTACCTCTCTTTTGGTGGTTCCTGTATCGTCTCAGATGATGACCTCCGGAGAT 701
Qy 540 TTCAAGTTTCAAGATGAAGAACGTGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 599
Db 702 TGCACCTTTTGGTCCCGAAATTCGAATTCAGTGTCTGTATGAGTTCATCCAGAAATAA 761
Qy 600 TGCTGCAATTTGGCTGTAGCCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 656
Db 762 GACGCTATTGTCGTTGTCAGTCAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 821
Qy 657 TGATGAGTCTCTTGTGGAGGCAATAGCCAAACAAA 692
Db 822 TGATGAGAAATATCTCGATGTTATCAGGAGACTAA 857

RESULT 5
US-08-181-629A-2
; Sequence 2, Application US/08181629A
; Patent No. 5472872
; GENERAL INFORMATION:
; APPLICANT: Swaminathan, Neela
; APPLICANT: Van Ethen, James
; APPLICANT: Mead, David
; APPLICANT: Skowron, Piotr
; TITLE OF INVENTION: Recombinant CviJI Restriction Endonuclease
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/181,629A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:

NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 31504
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5496 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-181-629A-2

Query Match 3.5%; Score 66.6; DB 1; Length 5496;
Best Local Similarity 52.7%; Pred. No. 2.5e-09;
Matches 144; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
Qy 978 TGGACACCCCTTGGCTTGGCTTCTTATGGCCACTTGGCTTGGCTTGGCTTGGCTTGGCT 1037
Db 3468 TTGGACGGCTTACATTTACCGCGCTTTTAAATGGTATAGCATGTGCGTCAAGACGCTAT 3527
Qy 1038 AGCCATGGTGTGATCTTGACAGCTTGGATGTCAAGGCACAGACGACCTTTTCACTGC 1097
Db 3528 TGATGGGGTGCAATCTTGACATACAGATATTTCCGGATGTACACCATTTCACTGC 3587
Qy 1098 TGTCACTGATGGCATCTGGACTGTGTACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1157
Db 3588 GGTATTATAGCACCATGATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3647
Qy 1158 TGTGATCATCTACAACTGT 1217
Db 3648 CGTCAITGATGATCTGAGTGGGTGCGTTACATTAACGCGCTTTTAAATGTAATGATGC 3707
Qy 1218 TATCTCGAGAGCTCTAGACCATGTGTGCGA 1250
Db 3708 GATTTGAGATGCTCATTAAGCAGGTGCAGA 3740

RESULT 6
US-09-509-802-1
; Sequence 1, Application US/09509802
; Patent No. 6489130
; GENERAL INFORMATION:
; APPLICANT: Immunex Corp.
; APPLICANT: Bird, Timothy
; APPLICANT: Virca, G.D.
; TITLE OF INVENTION: DEATH ASSOCIATED KINASE CONTAINING ANKYRIN REPEATS (DAKAR)
; FILE REFERENCE: 2889-US
; CURRENT APPLICATION NUMBER: US/09/509,802
; CURRENT FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 1
; LENGTH: 2370
; TYPE: DNA
; ORGANISM: Mus sp.
US-09-509-802-1

Query Match 3.2%; Score 60.2; DB 4; Length 2370;
Best Local Similarity 51.3%; Pred. No. 1.2e-07;
Matches 140; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
Qy 980 GGACACCCCTTGGCTTGGCTTCTTATGGCCACTTGGCTTGGCTTGGCTTGGCTTGGCT 1039
Db 1829 GGACACCCCTTGGCTTGGCTTCTTATGGCCACTTGGCTTGGCTTGGCTTGGCTTGGCT 1888
Qy 1040 CCATGT 1099
Db 1889 ACCTGTCTCTGTATGTTAATCTTGCAGCTTGCAGCTTGCAGCTTGCAGCTTGCAGCTTGC 1948